An *in silico* approach to identify novel genes preferentially expressed in the eye

Sapna P. Patel

PhD
University of Edinburgh
2004

Declaration

I declare:

(a) that I have composed this thesis

(b) that this work is my own

(c) that this work has not been submitted for any other degree or professional qualification

Sapna Patel

May 2004

Acknowledgements

I would like to thank Ian Jackson for his supervision and advice during the writing of my thesis. Also, thanks to Ian Smyth for his additional suggestions. I am grateful to Paul Perry, who gave up his spare time to help me when a Mac deserted me and took my images with it. Thanks to Alison, Dan and Eva- they saved me from nights on the cold streets of Edinburgh.

I would like to thank the people that gave me help and advice during the course of my PhD. Thanks to everyone at MRC Technology, particularly Gwen Cranston and Lisa Duffy who saw me through the highs and lows of *in situs*. Also, thanks to everyone in the Jackson group for general advice and help, particularly Lisa McKie who helped me out when I needed to be in two places at the same time.

Thanks to all the people who made the more important things in life fun. Thanks to Jo for the 5K and thanks to all the students for their good company- cheers Sue and Gill for making my dream of a 70s party come true. A big thank you to my parents for their support throughout and before my years in Edinburgh and to Duncan for support, advice and, most importantly, for making me laugh.

Finally, thanks to Jack Bauer. For a while there I was living from one 24 episode to the next- sad but true…

Abstract

Specific or preferential expression of a gene in a particular tissue or organ is often indicative of its importance in function or development. It would be beneficial to discover novel genes with preferential expression to enhance understanding of the tissue or organ in question. In this thesis, an *in silico* method was employed to find novel genes preferentially expressed in the mouse eye.

The chosen *in silico* method was digital differential display (DDD), which utilises the publicly available UniGene database as a resource. UniGene is composed of expressed sequence tag (EST) sequences that are clustered by homology, such that each cluster is considered to represent a gene. Crucially, ESTs in UniGene are traceable to their library of origin. This allows DDD to statistically compare, for each cluster, the proportional representation of ESTs in different user-defined pools of libraries. Using an eye pool and a non-eye pool of libraries, DDD was performed on the mouse, human and rat UniGene sets.

The DDD results used for further analysis comprised 292 mouse UniGene clusters that were more highly represented in the eye pool than in the non-eye pool. Of the 154 clusters that represented known genes, 48% were known to be eye relevant. Thirty-one of the 138 novel clusters were selected for expression studies. Ten of these were later recognised as known genes, with five having eye relevance, that is, previous evidence of preferential expression in the eye by *in situ* methods or mutations associated with eye defects. Fifteen of the 31 novel clusters were confirmed by RT-PCR to show preferential expression, with four appearing eye specific. Four of the 15 genes are known (*Aipl1*, *Crygf*, *Guca1b* and *Otx2*) and another nine putative genes show expression in various regions of the eye. These novel genes may well have a role in eye development or function and are candidates for mouse and human mutations that cause eye disease.

Contents

Supplementary Material

As well as the written information provided in this thesis, a CD of HTML pages is also provided. It contains additional information as well as a framework for easy access to relevant webpages and data. To open the homepage from which other information can be easily accessed, open the file "home.html". These pages are best viewed using Microsoft Internet Explorer.

List of figures

List of tables

Abbreviations

μ	micro
AIPL1	aryl hydrocarbon receptor-interacting protein-like 1
Amp	ampicillin
AMV	avian myeoblastosis
BCIP	5-bromo-4-chloro-3-indolyl phosphate
BLAST	basic local alignment search tool
bp	base pairs
BPAG1	bullous pemphigoid antigen 1
BRET	BTB/POZ gene, retinal
BTB	Broad-Complex, Tramtrack and Bric a brac
cDNA	complementary DNA
CDS	complete DNA sequence
CLN8	ceroid-lipofuscinosis, neuronal 8
CLN8L	ceroid-lipofuscinosis, neuronal 8-like
CNS	central nervous system
CRYGF	gamma crystallin F
D15Ertd806e	DNA segment, Chr 15, ERATO Doi 806, expressed
D5Bwg0860e	DNA segment, Chr 5, Brigham & Women's Genetics 0860 expressed
dATP	deoxyadenosine triphosphate
DCT	dopachrome tautomerase
dCTP	deoxycytidine triphosphate
DD	differential display
DDD	digital differential display
DEPC	diethyl pyrocarbonate
dGTP	deoxyguanosine triphosphate
dH2O	distilled water
DIG	digoxigenin
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleoside triphosphate
DST	dystonin
DTT	dithiothreitol
dTTP	deoxythymidine triphosphate
DUF676	domain of unknown function 676 protein

E	embryonic day
EB	elution buffer
EDTA	ethylenediaminetetraacetic acid
EST	expressed sequence tag
g	gramme
GABT4	gamma-aminobutyric acid (GABA-A) transporter 4
GAPD	glyceraldehyde-3-phosphate dehydrogenase
GCL	ganglion cell layer
GRTP1	GH regulated TBC protein 1
GUCA1B	guanylate cyclase activator 1B
HISS	heat-inactivated sheep serum
Hs	*Homo sapiens*
HTC	high throughput cDNA
IMAGE	Integrated Molecular Analysis of Genomes and their Expression
INL	inner nuclear layer
IPE	iris pigment epithelium
IPL	inner plexiform layer
IS	inner segments
kb	kilobase
LAG	longevity assurance gene
LFC	lens fibre cells
LGALS3BP	lectin, galactosidase-binding, soluble, 3 binding protein
m	milli
M	molar
MGI	Mouse Genome Informatics
Mm	*Mus musculus*
mnd	motor neuron disease mutation
mol	moles
mRNA	messenger ribonucleic acid
n	nano
N,N-DMF	N, N- dimethyl formamide
NAP	nucleosome assembly protein
NBT	nitroblue tetrazolium
NDR4	N-myc downstream regulated 4
NTMT	NaCl, TrisHCl, MgCl2, Tween
NTP	nucleotide triphosphate

OCT	optical cutting temperature
OFL	optic fibre layer
OLM	outer limiting membrane
OMIM	Online Mendelian Inheritance in Man
ONL	outer nuclear layer
OPL	outer plexiform layer
ORF	open reading frame
OS	outer segments
OTX2	orthodenticle homologue 2 (Drosophila)
p	pico
P	postnatal day
PBS	phosphate buffered saline
PBST	phosphate buffered saline with Tween
PCR	polymerase chain reaction
PFA	paraformaldehyde
POZ	pox virus and zinc finger
PPICAP	peptidylprolyl isomerase C-associated protein
PSAP	prosaposin
RDH12	retinol dehydrogenase 12
Rn	Rattus norvegicus
RNA	ribonucleic acid
RPE	retinal pigment epithelium
rpm	revolutions per minute
RT	reverse transcription
rUTP	radioactive uridine triphosphate
SAGE	serial analysis of gene expression
SAPAN	saposin A domain, novel
SET	suppressor of variegation, enhancer of zeste and Trithorax
SFTPB	surfactant associated protein B
SIX3	sine oculis-related homeobox 3 homologue (Drosophila)
SMART	simple modular architecture research tool
SSC	saline sodium citrate
SSH	suppression subtractive hybridisation
STGD3	stargardt disease 3

TBE	Tris, boric acid, EDTA
TBST	Tris buffered saline with Tween
TC	tentative consensus
TE	Tris, EDTA
TEA	triethanolamine
THC	tentative human consensus
TIGR	The Institute for Genomic Research
TLC	TRAM, LAG, CLN8
TRALE	transmembrane protein, retina and antisense lens expressed
TRAM	translocating chain-associating membrane protein
tRNA	transfer ribonucleic acid
Tsd	antisense to *Dst*
TSET	TSPY/SET related
TSPY	testis-specific protein, y-encoded
U	unit
UTR	untranslated region
VISTA	visualisation tools for alignments
X-Gal	5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside

CHAPTER 1

Introduction

1.1 The mammalian eye

The complexity of the eye has made it a much discussed and studied organ. In 'The Origin of Species', Darwin devoted a chapter to the proposal of theories that could explain its evolution by natural selection. However, the differing morphology of solutions to the light-sensing problem that are displayed by various organisms had since been considered evidence for their separate origins. More recently, evidence for the existence of a conserved genetic basis for development of the two very different types of eye found in insects and vertebrates was discovered in the homology of vertebrate *Pax6* (paired box gene 6) with *Drosophila eyeless* (Quiring et al., 1994). Subsequently, *Pax6* was found to induce ectopic eye formation in *Drosophila* (Halder et al., 1995), confirming its status as a master control gene and laying to rest doubts surrounding the common ancestry of these two different solutions to sensing light.

In addition to the interesting evolutionary aspect of the eye, the presence of many genetic diseases affecting this organ make research more directly applicable to human health. OMIM (Online Mendelian Inheritance in Man) has more than 200 diseases that include a form of retinal dystrophy (Sinha et al., 2000). In addition to retinal disease, cataract is a major worldwide cause of blindness, with glaucoma and corneal defects contributing to the diseases that cause visual impairment.

Although there are minor differences between the human and mouse eye, the mouse serves as the best available model for human eye disease. This is supported by a well-defined mouse genome that allows comparison to that of the human. Many mouse models of human eye diseases exist and some have been used successfully to locate the homologous human gene.

1.1.1 Gross eye structure

The mammalian eye (Figure 1.1) is composed of highly differentiated tissues that perform specialised functions, with light entering it being focused onto a layer of light sensitive receptors which convert photons to nerve impulses that are relayed to the brain and processed to form an image. At the core of the eye, there is a space filled with a fluid known as vitreous humour that maintains intraocular pressure, thereby supporting the shape of the eye. Three distinct layers surround this space, the retina being the innermost, the sclera the outermost protective layer and the choroid between the two, containing the vessels that supply the eye with blood. These layers surround most of the eye surface, with the exception of the anterior. Here, the transparent corneal layer covers an aperture, with aqueous humour filling the space between them and performing the same shape maintenance function as vitreous humour. Controlling the size of this gap, known as the pupil, is the iris, with the lens positioned posteriorly to these.

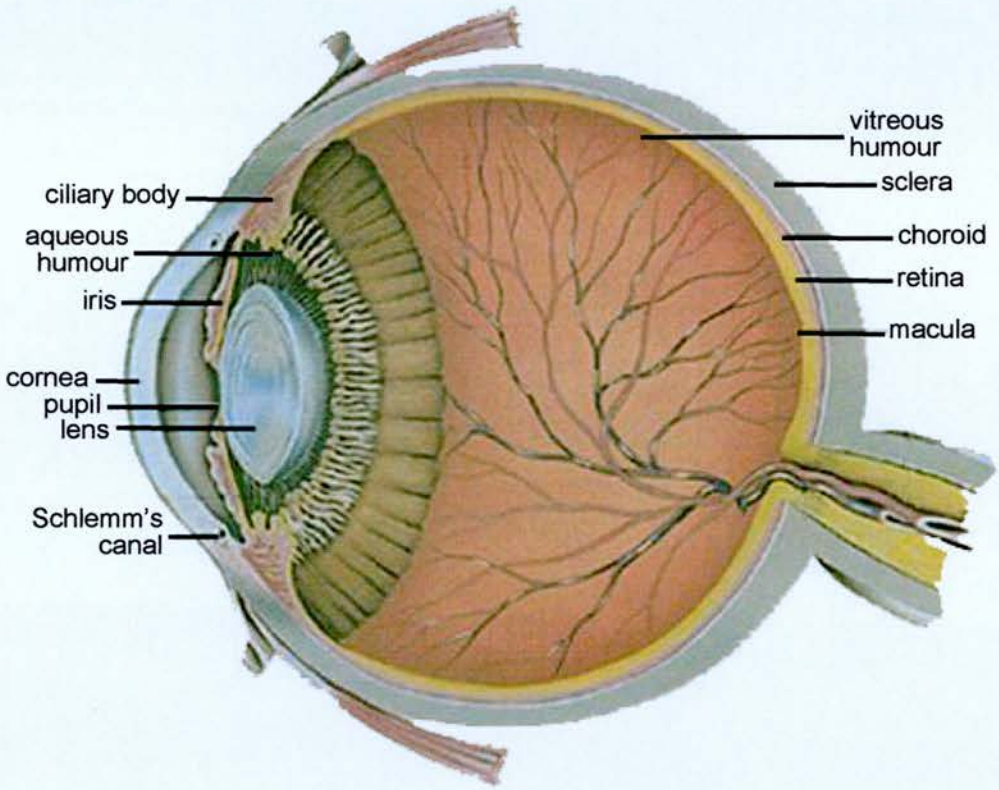


Figure 1.1 Structure of the human eye
The diagram presents a sagittal view of the human retina. The cornea acts as a protective barrier to the eye. Aqueous humour is secreted by the ciliary body and is filtered through Schlemm's canal before leaving the eye. The iris alters pupil size to control the amount of light that enters the transparent lens and passes through the vitreous humour. Light is focused onto the retina, containing a region known as the macula that is necessary for high visual acuity. The choroid, delivering blood to the eye, and the sclera form the outer layers. Adapted from an image taken from http://www.discoveryfund.org/anatomyoftheeye.html.

1.1.2 The cornea, iris and ciliary epithelium

The cornea consists of five layers, with the corneal epithelium as the most peripheral, and the other layers, moving towards the central eye, are Bowman's membrane, the stroma, Descemet's membrane and the endothelium. Corneal transparency allows light to enter the eye and pass through the aqueous humour, a transparent fluid contained in a chamber that lies behind the cornea.

Light that traverses the cornea and aqueous humour can then pass through the pupil. This is an opening that has its size, and therefore the amount of light passing through it, controlled by the iris. This pigmented structure, which also reduces glare in the eye, is muscular, with contraction and relaxation altering pupil size. Roles in controlling intraocular pressure (Chang et al., 1999) and regulating lens growth (McAvoy et al., 1999) have also been postulated.

The ciliary epithelium is a continuation of the iris epithelium, with a bilayer of pigmented and non-pigmented opposing cells (Raviola and Raviola, 1978). It secretes aqueous humour and therefore contributes to the regulation of intraocular pressure, as does the trabecular meshwork, a region through which the aqueous humour is filtered before exiting the eye through a structure known as Schlemm's canal.

1.1.3 The lens

Light passing through the pupil next encounters the lens (Figure 1.2). This structure has attached suspensory ligaments that control lens thickness and therefore control the refraction of light entering it, making it important for visual acuity. The lens is composed of anterior lens cells and lens fibre cells (LFCs). The anterior epithelial cells elongate and terminally differentiate into the LFCs, which are morphologically distinct, losing their nuclei and organelles. Lens transparency is maintained by this lack of internal structures, which would otherwise cause light to scatter, and also by the structural arrangement of its proteins. The anterior cells continue to differentiate in adulthood, although at a slower rate, thus the lens continues to grow throughout life.

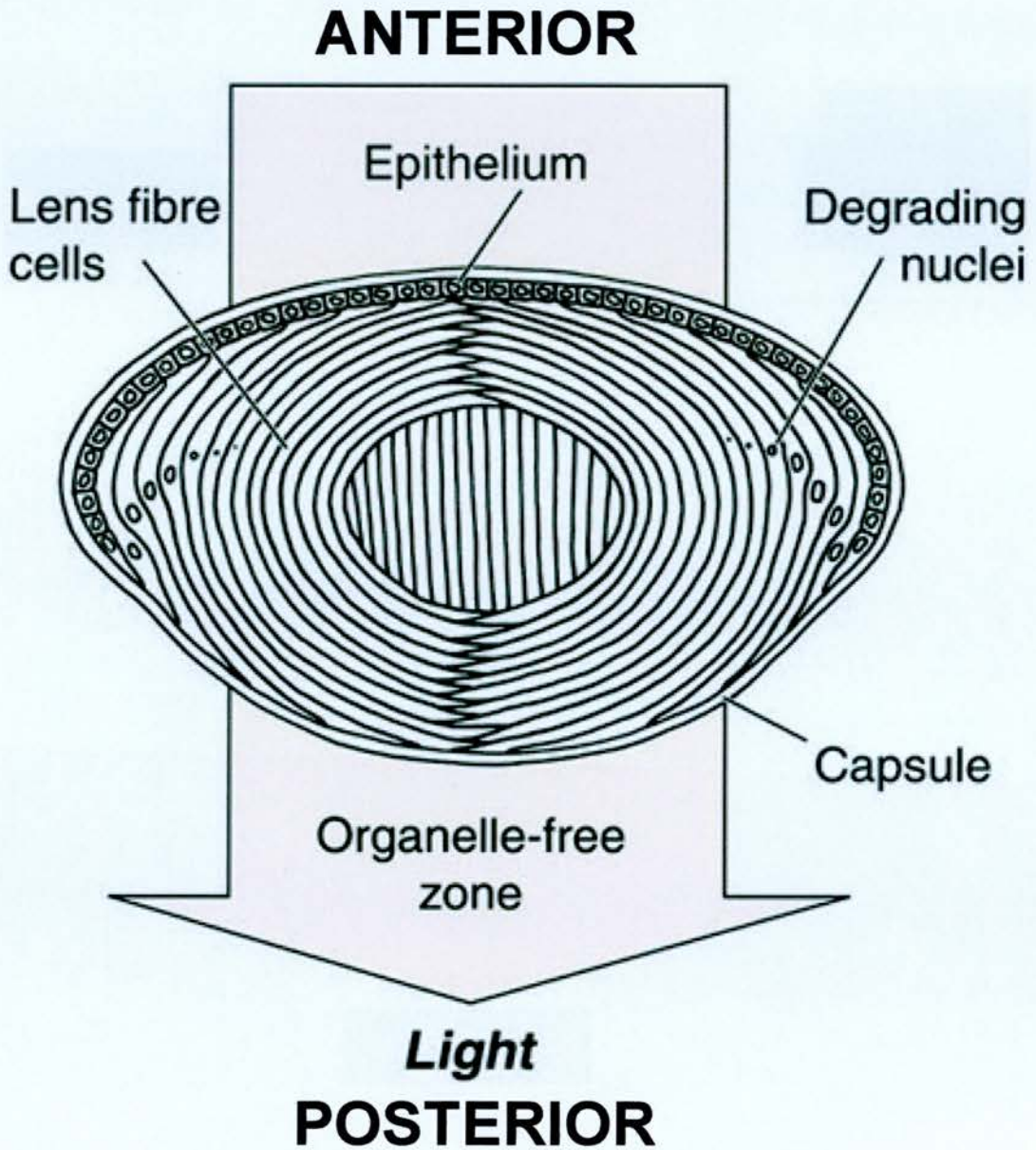


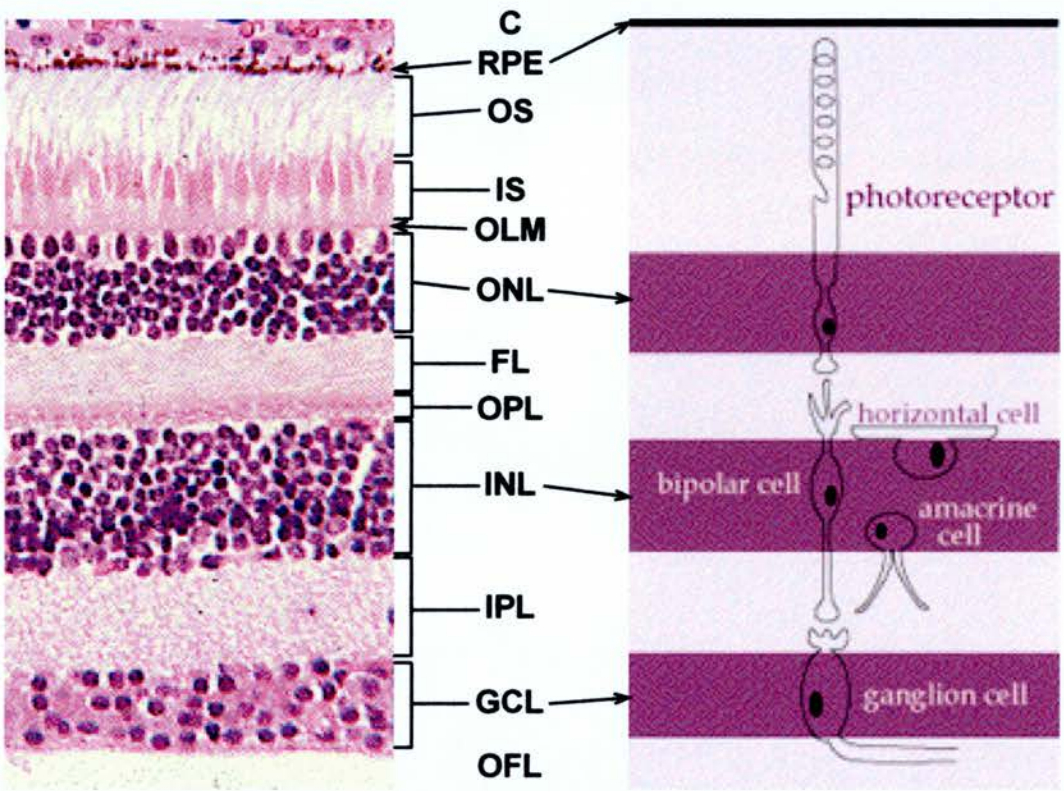
Figure 1.2 Lens cellular structure
The lens is composed of anterior epithelial cells and lens fibre cells that extend from the anterior cells to the posterior lens. Transparency is maintained by the lack of organelles present in the light path. Adapted from an image taken from http://www.eb.tuebingen.mpg.de/dept3/research_interests/eye_lens_dev.html.

1.1.4 The retina

Light passing through the lens is focused on the retina. This complex sensory tissue constitutes the innermost of three major layers on the outer surface of the eye. The retina contains the photoreceptive cells and neurones that transmit signals to the brain, which then produces an image.

The retina is in itself composed of multiple layers (Figure 1.3). Separating the retina from the vitreous humour is the inner limiting membrane. Adjacent to this, the innermost layers are the optic fibre layer (OFL) and the ganglion cell layer (GCL), which contain ganglion cells and some amacrine cells. The next three layers are the inner plexiform layer (IPL), the inner nuclear layer (INL) and the outer plexiform layer (OPL). These layers incorporate ganglion, bipolar, horizontal and amacrine neuronal cell types. Beyond these cells lie another fibre layer and the outer nuclear layer (ONL). The inner and outer segments of the rod and cone photoreceptor cells form the next two layers (IS and OS), while the outer limiting membrane (OLM) separates these from the ONL. Müller glial cells span the layers of the retina from the OFL to the inner segments of the rods and cones. These layers of neuronal and glial cells constitute the neural retina. The surface of the retina is not uniform in its cellular composition. This is most obviously illustrated by the human fovea, a small area of the neural retina diametrically opposite the cornea and pupil that contains only cone cells. The fovea lies in the macula, 6mm in diameter in primate retinas, which has a high proportion of cones, in contrast to the peripheral retina. Rods are necessary for vision in dim light while cones are important in bright light conditions and for colour vision and visual acuity. In the mouse, cones are much more sparse than in the human retina and the fovea does not exist.

PERIPHERAL EYE



INNER EYE

Figure 1.3 Cellular composition of the retina
The multiple layers of the retina are shown alongside their corresponding cell types. The Muller cells (not shown) span the layers from the OFL to the IS. Ganglion cells are also found in more peripheral layers and amacrine cells are found in the GCL. As well as the retina, the choroid is shown. The inner limiting membrane (not shown) marks the inner boundary of the retina, separating the vitreous humour from the OFL. C, choroid; RPE, retinal pigment epithelium; OS, outer segments; IS, inner segments; OLM, outer limiting membrane; ONL, outer nuclear layer; FL, fibre layer; OPL, outer plexiform layer; INL, inner nuclear layer; IPL, inner plexiform layer; GCL ganglion cell layer; OFL, optic fibre layer. Adapted from an image taken from http://thalamus.wustl.edu/course/eyeret.html.

The retinal pigment epithelium (RPE) is the outermost layer of the retina, separated from the choroid by Bruch's membrane. It is pigmented, as the iris and ciliary epithelia are, and is continuous with these tissues. The RPE contributes to the maintenance of the neural retina, being involved in selective transport of nutrients and ions, synthesis of the photoreceptor extracellular matrix, phagocytosis of shed outer segments and maintenance of the blood/neural retina barrier (Bok, 1985; Bok, 1993). It also has a major role in the visual cycle (see section 1.1.5).

1.1.5 Phototransduction and the visual cycle

There are two phases involved in phototransduction. The excitation phase describes the processes leading to hyperpolarisation and the production of a stimulus, while the termination of transduction and recycling of its components occur in the recovery phase.

Rhodopsin is the light sensitive pigment contained in rod photoreceptors. Photoactivation of rhodopsin initiates the excitation phase of phototransduction (Figure 1.4). This alters the conformation of transducin, a G protein, by converting the GDP to which it is bound to α-GTP. The inhibitory γ subunit of phosphodiesterase (PDE) is removed by interaction with the transducin-bound α-GTP. The subsequently active form of PDE hydrolyses cGMP present in the cytoplasm. There are cGMP-gated channels in the membrane of the outer segments, thus as the concentration of cGMP decreases intracellularly, the channels close. Transport of calcium and sodium into the outer segments decreases, causing hyperpolarisation. (Lamb, 1996)

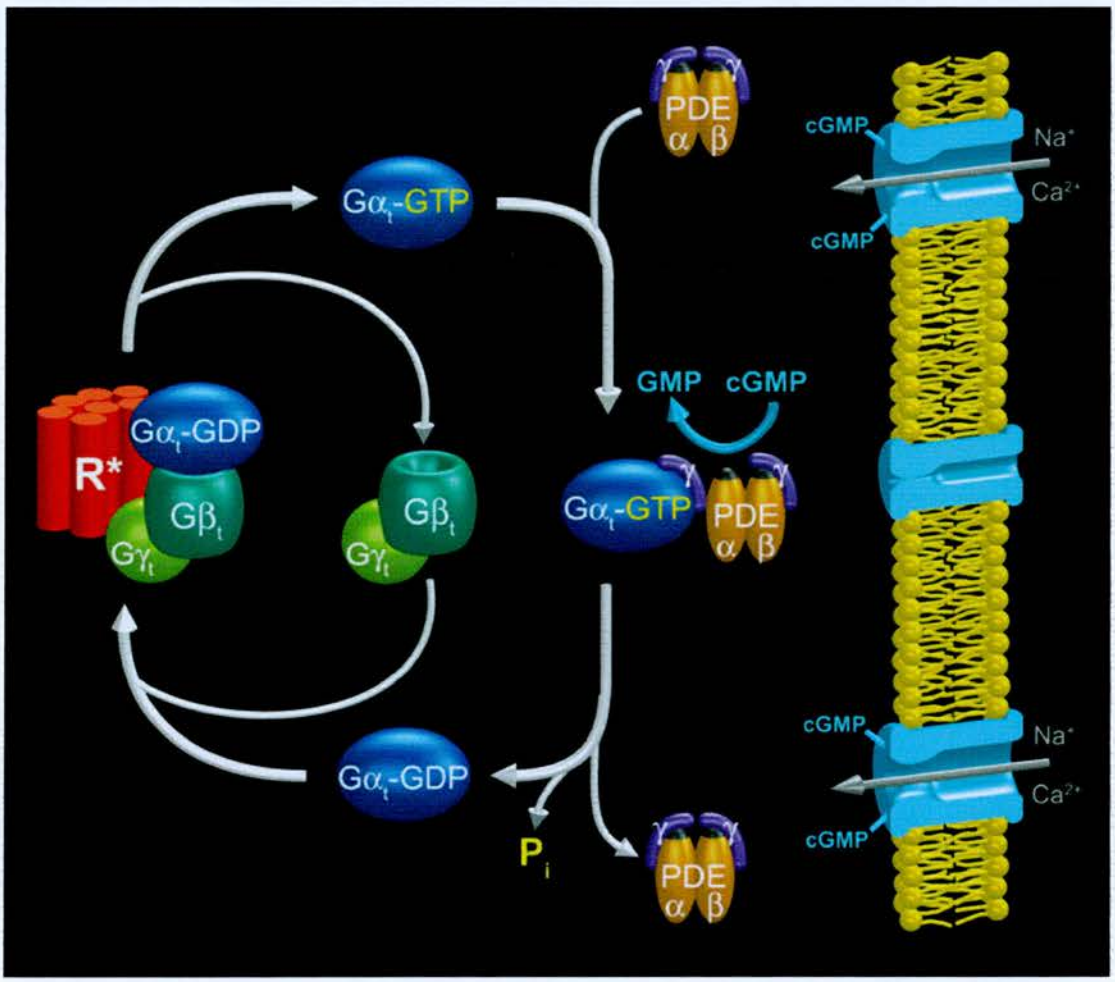


Figure 1.4 The phototransduction cascade
Photoactivated rhodopsin alters the conformation of the bound G protein, transducin. The α-GDP subunit is converted to an unbound α-GTP. This can then bind the γ-PDE subunit, allowing PDE activation. This activity converts cGMP, which gates calcium and sodium channels in the cytoplasm, to GMP. Closed channels result in a decrease in intracellular ions and causes hyperpolarisation. R*, active rhodopsin; t, transducin; G, G protein; GMP, -DP, -TP, guanosine -mono, -di, -triphosphate; PDE, phosphodiesterase. Taken from http://www.howelaboratory.harvard.edu/vahomepage/research_arshavsky.htm.

The recovery phase sees the return of rhodopsin, transducin and PDE to their inactive states. Rhodopsin is recovered using rhodopsin kinase and arrestin via the visual cycle (described below), while a G protein mediates the hydrolysis of transducin-associated GTP to GDP. In turn, GDP inactivates PDE. Activated guanylate cyclase catalyses the increase of cGMP and the channels open.

The visual cycle (Figure 1.5) takes place in the RPE and regenerates photoreceptor pigment. In its inactive state, rhodopsin is composed of opsin protein and 11-*cis* retinal. Photoactivation of rhodopsin changes the bound 11-*cis* retinal to all-*trans* retinal. The regeneration of inactive rhodopsin therefore necessitates the conversion of all-*trans* retinal to 11-*cis* retinal, which is the essence of the visual cycle. This process is characterised by a series of chemical modifications, most of which take place in the RPE, with an influx of the all-trans retinol component (an alcohol derivative of vitamin A) obtained from the choroidal blood supply.

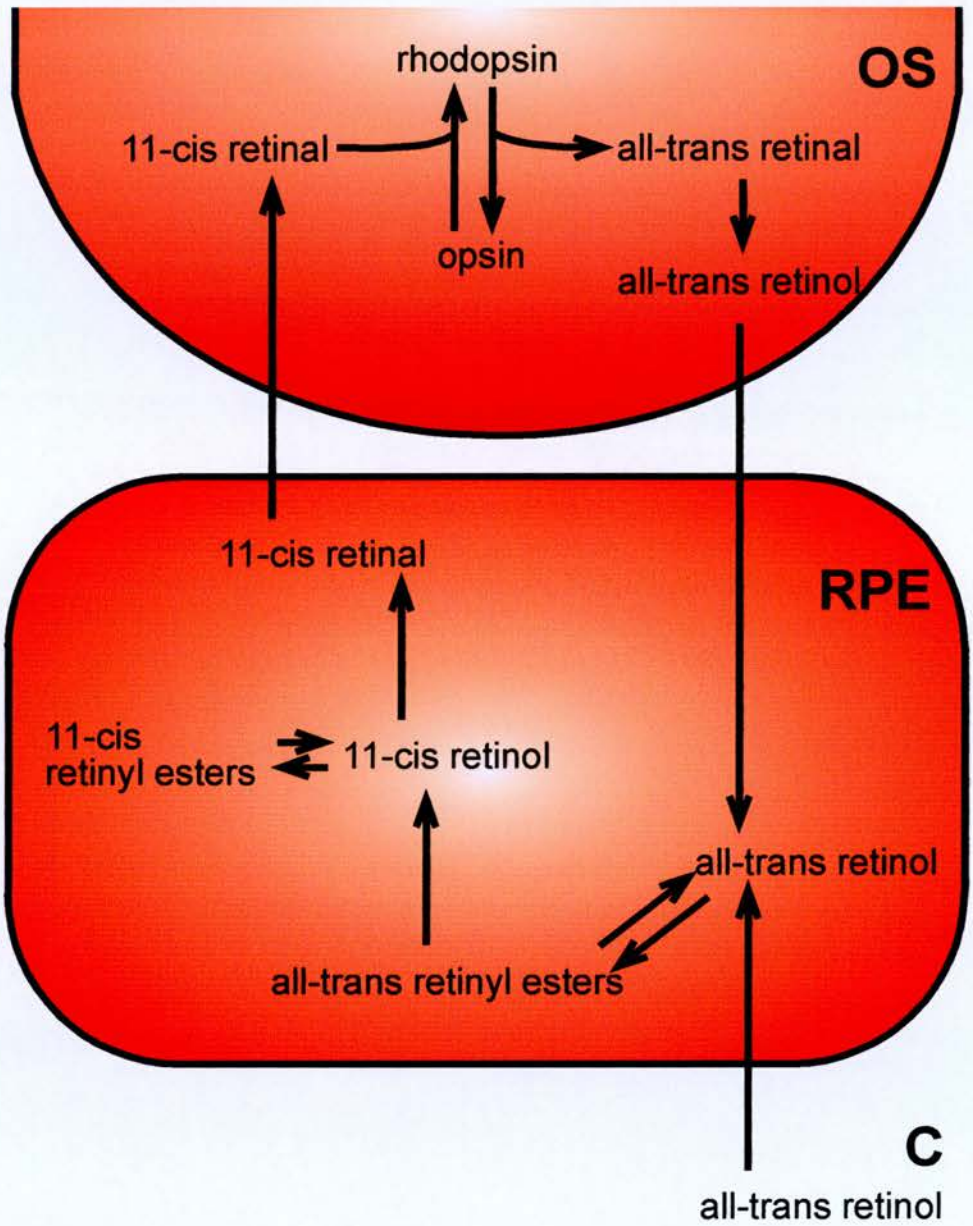


Figure 1.5 The visual cycle
The visual cycle describes the regeneration of rhodopsin. Photoactivation of rhodopsin changes the opsin-bound 11-*cis* retinal to all-*trans* retinal, with the regeneration of rhodopsin brought about by the conversion of all-*trans* retinal back to 11-*cis* retinal. Initially, the released all-*trans* retinal is reduced to all-*trans* retinol. This is transported out of the OS into the RPE, where all-*trans* retinol, an alcohol derivative of vitamin A, can also be acquired from the bloodstream. Esterification of all-*trans* retinol is followed by isomerisation. Hydrolysis and subsequent oxidation result in 11-*cis* retinal, which is transported from the RPE into the OS to bind with opsin and reform active rhodopsin. OS, outer segment; RPE, retinal pigment epithelium; C, choroid. Based on a figure taken from Rattner et al., 1999.

1.2 Approaches to gene discovery

Novel genes can be found through various methods. A top down approach may be the investigation of an interesting phenotype to elucidate the gene or genes that may be genotypically responsible. This would culminate in the mapping of a particular mutation, which has historically been spontaneous. ENU (ethylnitrosonurea) mutagenesis induces point mutations chemically and was first employed to screen for specific mutations in the progeny of mutagenised mice in 1984 (Bode, 1984). Gene trap methods have also been used and involve inserting a reporter gene into an endogenous ES cell gene, which not only removes function of the gene of interest, but also allows the visualisation of the expression domain of that gene through transcription of the reporter gene by the endogenous promoter (Gridley et al., 1987; Jaenisch, 1988). These methods for gene discovery, like many others, are based on the premise that the laboratory mouse is the best existing model for the human (Cox and Brown, 2003).

The comparison of mouse and human is made more direct by the advent of well-described genomes for these organisms. Related to this change is the possibility of using bottom up approaches to gene discovery, by locating a novel gene and using experimental techniques, such as expression studies and experimentally altering genotype in model organisms to give a phenotypic effect that can be related to function. Novel genes can now be located using features in the human genome such as CpG islands, splice sites and open reading frames (ORFs) and the use of tools such as BLAST allow orthologous and paralogous genes to be discovered (Cox and Brown, 2003).

1.2.1 The advent of ESTs

The potential for identifying genes from cDNA sequences was first realised in 1985, with the publication of a novel gene found by shotgun sequencing (Putney et al., 1983). The advent of fluorescent-based sequencing (Smith et al., 1986) allowed an automated method to develop and be used in a more high throughput manner. The term EST (expressed sequence tag) was coined by Venter's group in the early 90s and describes a short sequence, usually between 200 and 400 bases in length, with approximately 2% error, derived from a cDNA clone. These sequences were first produced and used in a pilot study using brain cDNA libraries (Adams et al., 1991).

Adams *et al.* had recognised the potential of ESTs to identify genes and its relative rapidity and low cost compared to sequencing the entire genome. Sequencing ESTs has been described as the hare to the human genome project (HGP) tortoise (Gibbs, 1997), as while ESTs are a more efficient and direct approach to attain a complete set of genes by targeting the coding 2% of the genome, the HGP provides the framework on which to place these coding genes and the peripheral sequences that affect their expression and function.

Given the advantages, EST sequencing was generally accepted as a good complementary approach to genomic sequencing and therefore prompted the production of thousands of ESTs by 1993. In order to provide these sequences in a publicly accessible form, GenBank (Benson et al., 2003), the sequence submission database, set up an EST arm known as dbEST (Boguski et al., 1993).

Large-scale studies of ESTs from the human genome were undertaken and have contributed large amounts of data to dbEST. TIGR (The Institute for Genomic Research) is an organisation that was set up for the purpose of sequence-based research. The automated DNA sequencing and analysis equipment present at this centre have enabled the high throughput sequencing that facilitates large-scale EST

analysis. The initial result of this activity was the production of over 170,000 ESTs from 300 cDNA libraries encompassing 37 different tissues and organs (Adams et al., 1995). Subsequent major EST producers have used human (Hillier et al., 1996; Houlgatte et al., 1995) and mouse (Marra et al., 1999) libraries.

IMAGE (Integrated Molecular Analysis of Genomes and their Expression) is a consortium that was founded in an effort to pool cDNA libraries in a public database and provide a single identifier for each, in order to provide expression data on a single clone that may have been gathered by many different laboratories (Lennon et al., 1996).

ESTs are still the greatest source of new sequence data submitted to GenBank, with human and mouse sequences constituting the largest numbers (Benson et al., 2003).

1.2.2 Using ESTs to find novel genes

By 1994, ESTs were already fulfilling their predicted function as a tool for gene discovery (Sikela and Auffray, 1993; Boguski et al., 1994). However, the efficiency of using ESTs to extract novel sequences was severely hampered by the large variation in frequency for different mRNAs. This ranges from one copy per cell to an average of 5,000 copies amongst those mRNAs contributing 40% to 45% of the total population (Davidson and Britten, 1979; Bishop et al., 1974). In order to reduce this inherent problem, a normalisation technique was applied to the construction of a brain cDNA library to allow representation to fall within a narrower range (Soares et al., 1994). This technique relies on reassocation kinetics, with rarer transcripts annealing less rapidly than common ones, and bound sequences being removed in the procedure (Figure 1.6). As well as allowing rare transcripts to be more readily identified from cDNA libraries, the cost and time for sequencing ESTs was reduced and has allowed the evolution of the large-scale analysis that is now carried out,

including a recent "cDNA encyclopaedia" of the mouse that represents 70,000 transcriptional units (Carninci et al., 2003).

A further advancement in EST studies was made through the presentation of ESTs for analysis. Although dbEST serves the crucial purpose of allowing public access to ESTs, there was an additional need for a more organised and structured EST database, grouping ESTs from common transcripts and allowing more direct comparison with the ongoing annotation of the human genome. Two similar but distinct databases were set up, both providing a set of clusters containing ESTs that had been grouped by homology. At TIGR, tentative human consensus (THC) sequences were produced from homologous ESTs (Adams et al., 1995; Quackenbush et al., 2000), while NCBI (National Center for Biotechnology Information) produced the UniGene database (Schuler et al., 1996; Pontius et al., 2003; Wheeler et al., 2003). UniGene comprises a set of clusters containing homologous ESTs and, in cases where these have been derived from known genes, mRNA sequences.

ESTs provide a vast amount of sequence information that can be used to find novel genes. Improved techniques, such as normalisation, and the provision of organisational EST databases, UniGene and THCs being the prime examples, have allowed this resource to be fully exploited.

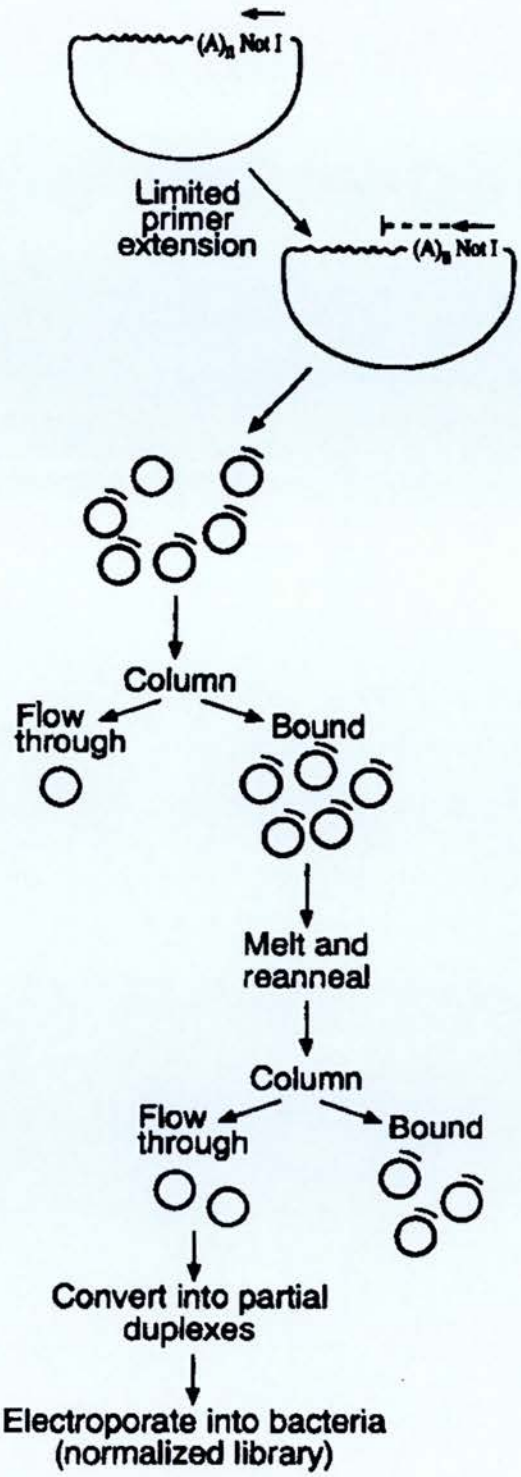


Figure 1.6 Schematic of cDNA library normalisation
The normalisation procedure aims to equalise the abundance of different cDNA species from a library in order to make sequencing more efficient and increase the likelihood of sequencing rare cDNAs. The first step requires primer extension to produce complementary sequences of cDNA clones of about 200 bases. The partial duplexes resulting from this are purified and subsequently melted and reannealed. Rarer sequences anneal less rapidly than common sequences, due to reassociation kinetics. The resulting duplexes, representing common sequences, are extracted to leave unhybridised cDNAs that are used to construct the normalised library. Adapted from a figure taken from Soares et al., 1994.

1.3 Differential expression

Expression analysis is used as a tool to discover genes involved in the development or function of a particular tissue or cell type. This is based on an assumption that differential expression of a gene implies an important role in that restricted group of cells or tissues. Application of this technique to all cell types and tissues may eventually lead to a "body map of expressed genes" (Okubo et al., 1992), an additional level of complexity to the genome sequence.

There are many methods for assessing gene expression and discovering genes with differential expression. This section describes these techniques, their strengths and weaknesses and their previous application.

1.3.1 Northern blot analysis

Since the development of the Northern hybridisation technique (Alwine et al., 1977), it has been used to assess gene expression. The simplest form of this technique is the hybridisation of a radioactively labelled probe to RNA samples. The intensity of a signal on radiographic film is indicative of the level of RNA present. Relative concentrations can be deduced by using a dot blot technique, whereby multiple samples are dotted on a filter before hybridisation (Kafatos et al., 1979). This approach has many disadvantages compared to more recently developed techniques. There are issues with the reproducibility of results and therefore quantification of RNA can not be accurately assessed. It is also a technique that is unsuitable for large-scale analysis, as only one gene can be assessed at a time and the amount of RNA required for this type of approach is prohibitive to this type of study.

1.3.2 RT-PCR

Following the publication of the polymerase chain reaction (PCR) (Mullis et al., 1986), RT-PCR (reverse transcription-PCR) was soon developed and used to identify a mouse mutation causing sparse fur (Veres et al., 1987). This technique amplifies the relevant population of cDNAs derived from an RNA sample and therefore presents a relatively simple method for assessing differential expression of a gene. Various modifications of the basic technique have allowed quantification of cDNA, either by assessing quantity once the reaction has been completed, or more accurately by real-time PCR, monitoring the product as the reaction progresses (Freeman et al., 1999). A major advantage of RT-PCR over using Northern analysis is the small quantity of RNA that can be used for a reaction, as it much more sensitive. Additionally, a large number of samples can be tested at the same time. However, as with dot blots, the limiting factor is the number of genes that can be tested at one time, with optimised conditions for quantitative RT-PCR being specific for each primer set.

1.3.3 Sequencing ESTs from cDNA libraries

The generation of ESTs has already been discussed as a successful way of finding novel genes. In addition to this function, ESTs have been used to quantify gene representation in cDNA libraries that have not been normalised. Original studies merely aimed to build a profile of genes expressed in a tissue, with the first such study constructing a cDNA library from a liver cancer cell line (Okubo et al., 1992). A follow up investigation sequenced ESTs from liver and lung cell libraries and compared the expression profiles of these tissues with that deduced for the liver cancer cell line in the first study in order to classify genes as tissue specific or non-

specific (Matsubara and Okubo, 1993). This approach has since been applied to a range of tissues and cell types.

As EST sequencing does not require prior knowledge of a sequence, it presents the opportunity to discover novel genes with differential expression as well as establish tissue specificity for known genes, unlike other techniques for assessing gene representation. However, saturation of the transcript population of a cell can only be achieved with extensive sequencing. As previously discussed, this is inefficient, with common transcripts being sequenced repeatedly. However, inherent in the normalisation procedure is the altered representation of transcripts. Thus, sequencing from a single cDNA library constructed using the cells or tissue of interest has its limitations.

1.3.4 Subtractive hybridisation

As sequencing cDNA libraries is an inefficient process for gaining transcriptional profiles for cell or tissue types, subtraction is a method that can be employed to produce a population of differentially expressed sequences (Figure 1.7). The subtraction approach was first utilised to find T lymphocyte specific proteins by using RNA derived from B lymphocytes to hybridise, and therefore remove, non-specific T lymphocyte cDNAs (Hedrick et al., 1984). This method produces sequences that are preferentially expressed in one cell or tissue type (tester) and not another (driver) and allows the discovery of differentially expressed novel genes.

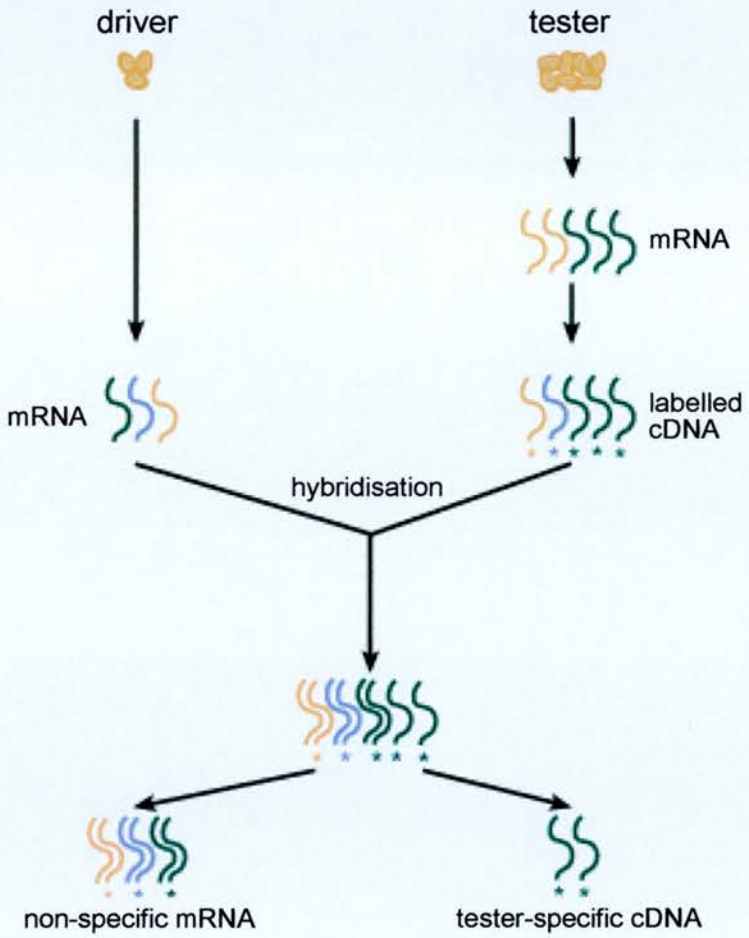


Figure 1.7 Schematic of subtractive hybridisation
Subtraction can be used to produce sequences that are preferentially expressed in a particular tissue or cell type. The mRNA of these cells, known as the tester, is used to synthesise labelled cDNA. Hybridisation between the tester cDNA population and the mRNA from a driver tissue result in a mixed population of hybridised and non-hybridised sequences. The hybridised sequences are removed to leave labelled sequences from the tester that represent cell or tissue specific transcripts. Adapted from an image taken from http://www.niaaa.nih.gov/gallery/genetics/outline.htm.

One of the disadvantages of subtraction is the high redundancy of clones. A modification of the subtraction method, suppression subtractive hybridisation (SSH), reduces this problem by combining subtraction and normalisation in one step (Diatchenko et al., 1996). SSH is based on suppression PCR, a technique for walking in uncloned genomic DNA from gene specific sequences (Ma and Whitlock, Jr., 1997). In SSH, tester cDNA is divided into two pools, each added to with distinct adapter sequences (Figure 1.8). The first hybridisation is with the driver in each pool, followed by a subsequent hybridisation of the two normalised tester pools with each other. Only sequences with both adapters will be amplified exponentially by PCR with adapter specific primers, so allowing the equalisation of cDNA representation whilst removing sequences not expressed differentially.

Although SSH removes the redundancy that occurs in standard subtraction hybridisation, there remains the disadvantage of only allowing the elucidation of upregulated expression in one particular sample.

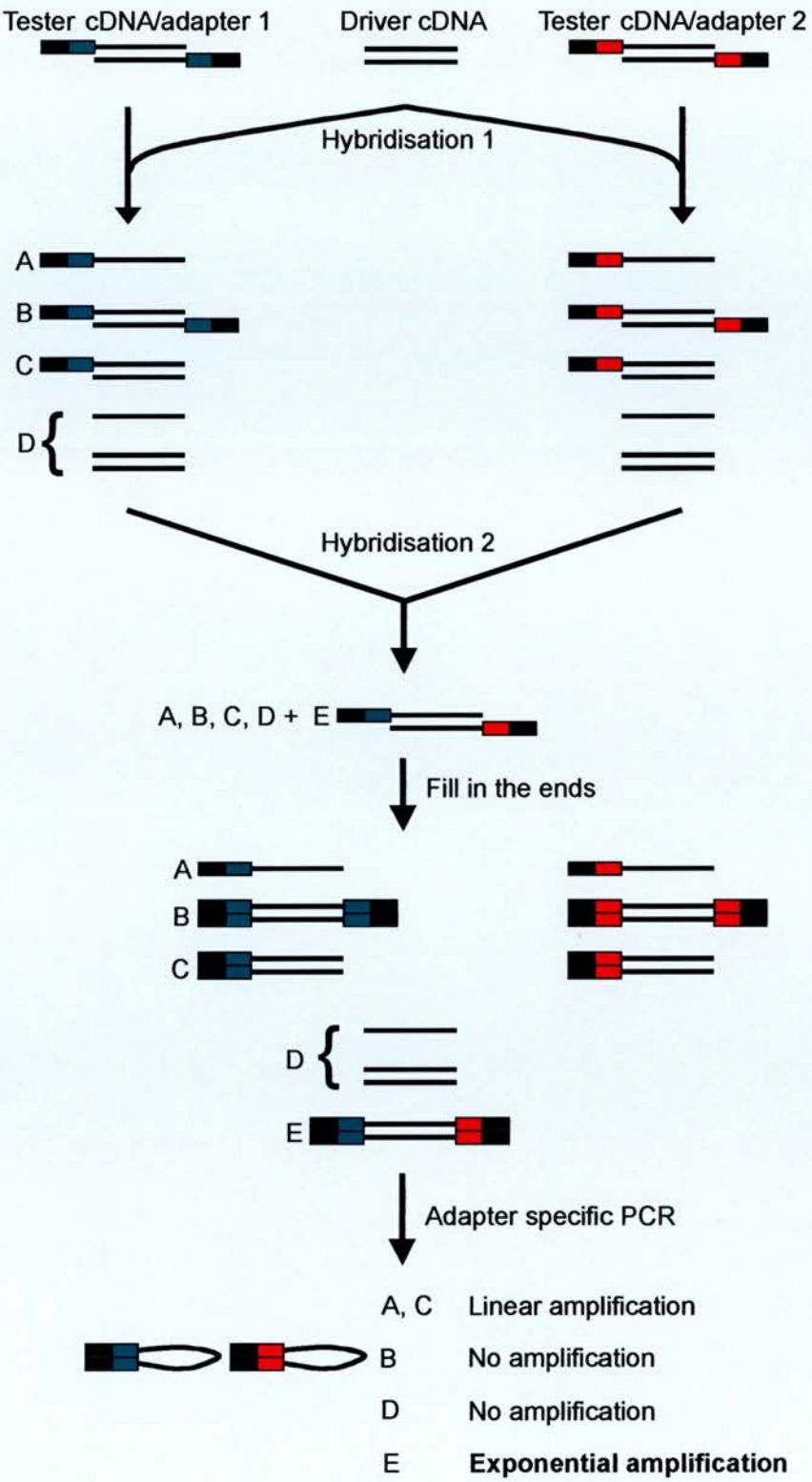


Figure 1.8 Schematic of suppression subtractive hybridisation
Two pools of cDNA derived from the tissue of interest (the tester) are ligated with different adapters. Each pool of sequences is hybridised to driver cDNA separately in the first hybridisation. The second hybridisation uses both tester/driver pools. The ends of products are then filled in and subjected to PCR using adapter specific primers. Only tester specific sequences that have not hybridised within the same tester pool are amplified by PCR.

1.3.5 *In situ* hybridisation

Since the first publication of wholemount *in situ* hybridisation (Herrmann, 1991), this technique has been widely used to visualise gene expression in a tissue or embryo. In order to use this tool as an indicator of differential expression, large-scale studies have been carried out to pinpoint differentially expressed genes, for example, in *Xenopus* (Gawantka et al., 1998) and mouse embryo (Neidhardt et al., 2000). In these studies, 1,765 and 5,348 cDNAs were analysed in *Xenopus* and mouse embryo respectively, yielding 273 and 317 differentially expressed genes. This approach can be used with novel sequences and may give a more detailed view of differential expression, as it can be seen *in situ*. However, it is a laborious method for screening large numbers of unbiased cDNAs to discover those that are tissue specific.

1.3.6 Differential display

Differential display (DD) is a method that allows visualisation of differences between expressed sequences across multiple RNA samples. Also known as DD-PCR, this technique requires the use of non-specific primers, amplifying a variety of sequences by PCR that can subsequently be separated by size electrophoretically. The presence or absence of these bands, and in some cases the intensity of a band, compared to the equivalent band in another sample can be used as an indication of up- or down-regulation of a gene (Figure 1.9). This was first developed and used to identify genes differentially expressed between a mammary epithelial cell line and a breast cancer cell line, with four sequences characterised and three of these confirmed as differentially expressed by Northern blotting (Liang and Pardee, 1992). Another group independently developed an almost identical protocol at the same

time and compared kidney to heart, characterising five sequences and confirming all to be differentially expressed (Welsh et al., 1992).

DD-PCR has since been a much used method for elucidating differentially expressed genes and has the advantages of allowing the visualisation of relative amounts of sequence representation across many samples, with no prior knowledge of sequences required for their investigation. However, this technique is difficult to automate since it requires visual comparison and removal of the relevant sequences for subsequent cloning.

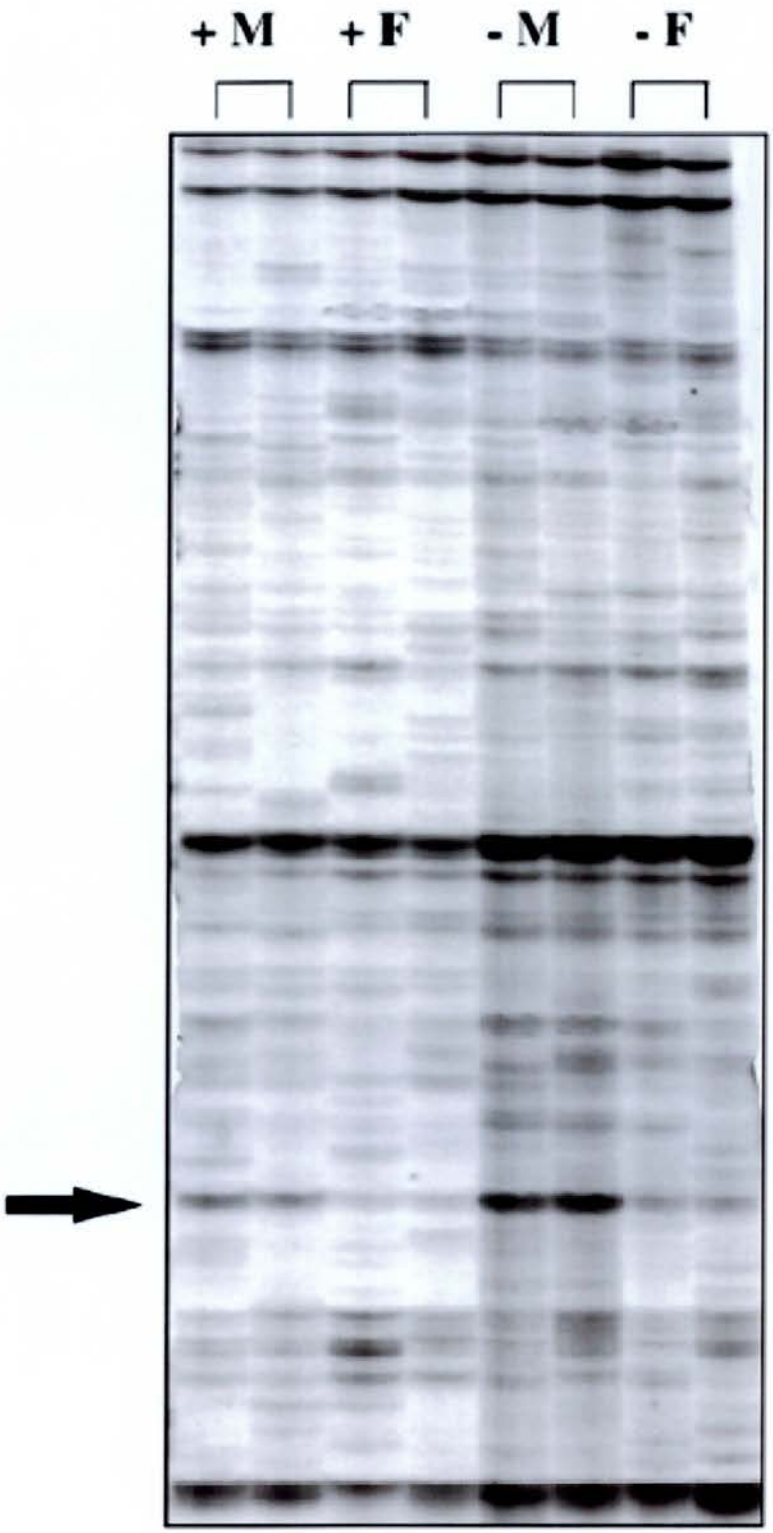


Figure 1.9 An example of differential display
Differential display allows the visualisation of randomly primed PCR products between different tissue or cell types. In this example, male (M) and female (F) rat lenses were either treated with a growth factor associated with cataract susceptibility (+) or untreated (-) and samples were run in duplicate. The arrow indicates a sequence that is differentially expressed between male and female and also appears to be downregulated in treated lenses. Taken from Sun et al., 2000.

1.3.7 SAGE

Serial Analysis of Gene Expression (SAGE) utilises a chain of biochemical steps to produce a series of short DNA fragments, each representing an mRNA, and therefore can produce sequences that represent the mRNA population of a given cell or tissue. This provides an alternative to sequencing ESTs derived from a specific source, which has been discussed previously (section 1.3.3). The SAGE technique (Figure 1.10) was first used in 1995 and describes the production and concatenation of these short 9bp cDNA fragments that can be used to identify a specific sequence (Velculescu et al., 1995). Thus, sequencing of these concatenations allows the quantification of a particular mRNA in the cell or tissue of interest. In this initial study, SAGE was applied to pancreas mRNA with success, as the ten most abundant tags corresponded to gene transcripts with defined pancreatic functions. Since its conception, SAGE has been used to compare expression profiles of various tissue and cell types. The comparison of many samples and the possibility of discovering novel genes are clear strengths of this approach. Additionally, the key advantage over EST sequencing lies in its lower cost. However, automated sequencing is required, as 30,000 to 100,000 tags need to be sequenced in order to make the results statistically viable (Fryer et al., 2002) and the construction of a SAGE library remains a costly and difficult procedure. SAGE is also limited to use in organisms with well-defined genomes, as sequences derived from these techniques can only then be compared with sequences already present in public databases.

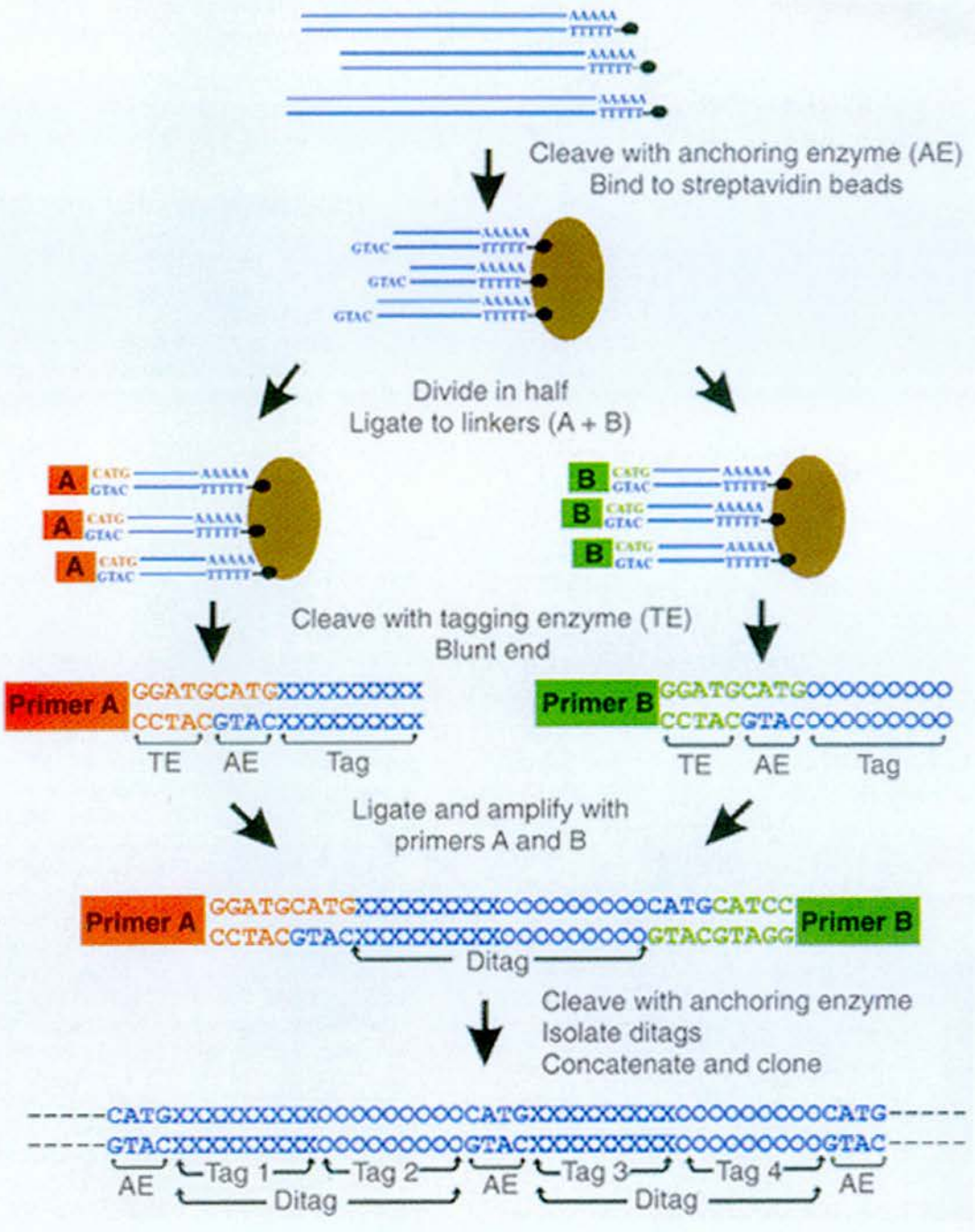


Figure 1.10 Schematic of SAGE
SAGE sequences and quantifies cDNA fragments representing transcripts from a particular cell or tissue. The mRNA is used to synthesise double-stranded cDNA which is subsequently anchored and restricted. Two pools are created and two different linkers are ligated to these. A restriction endonuclease is used to cleave the sequence at a specific distance from its site (TE), releasing sequences from their bound state. Ligation of these tags produces end-to-end sequences that can be cleaved to produce ditags. These ditags are concatenated and sequenced, with each tag identifying a specific sequence, allowing quantification of a transcript in a SAGE library. Taken from Velculescu et al., 1995.

1.3.8 Arrays

Array technology began with the use of what are now known as macroarrays (Zhao et al., 1995; Takahashi et al., 1995; Nguyen et al., 1995). This technique, also known as the reverse Northern, robotically blots different cDNA sequences representing genes onto a nylon membrane. These are referred to as the probes and the target used is RNA, which is placed onto the array to allow hybridisation. Radioactivity is used as the method of detection, with relative intensities used as indicators of the relative abundance (Figure 1.11). In this way, a large number of genes can be tested using one sample and the array can be used again with different RNA samples after washing. In initial studies using this technique, Northern blotting was used to validate the results.

Microarrays work on the same principle as macroarrays, using an array containing probes and hybridising a target sample to this array to gain an indication of relative abundance. However, microarrays have a larger number and higher density of probes on a glass or plastic surface and they utilise fluorescence rather than radioactivity as a measure of sequence abundance. An advantage of using fluorescence is the ability to distinguish between different fluorescence types, therefore more than one sample can be hybridised at one time due to competitive hybridisation (Figure 1.12). There are two major types of microarray that are based on different probe synthesis techniques; cDNA arrays and oligonucleotide arrays. The former consists of cDNAs that are placed on the array surface and was first used to successfully identify sequences differentially expressed between the root and leaf of *Arabidopsis* (Schena et al., 1995). Oligonucleotide arrays require *in situ* synthesis of 25 base probes on a silicon chip (Lipshutz et al., 1999). As the sequences are short, non-specific binding is an inherent problem, which is reduced by using a control sequence with a mismatch.

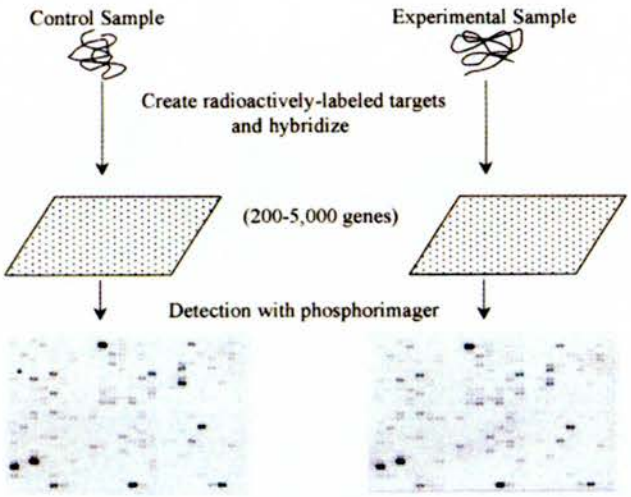


Figure 1.11 Macroarray procedure
RNA samples are hybridised to a membrane spotted with cDNA sequences representative of different genes. Multiple samples can be used for hybridisation to produce a visual display indicating relative sequence abundance. Adapted from a figure from Freeman et al., 2000.

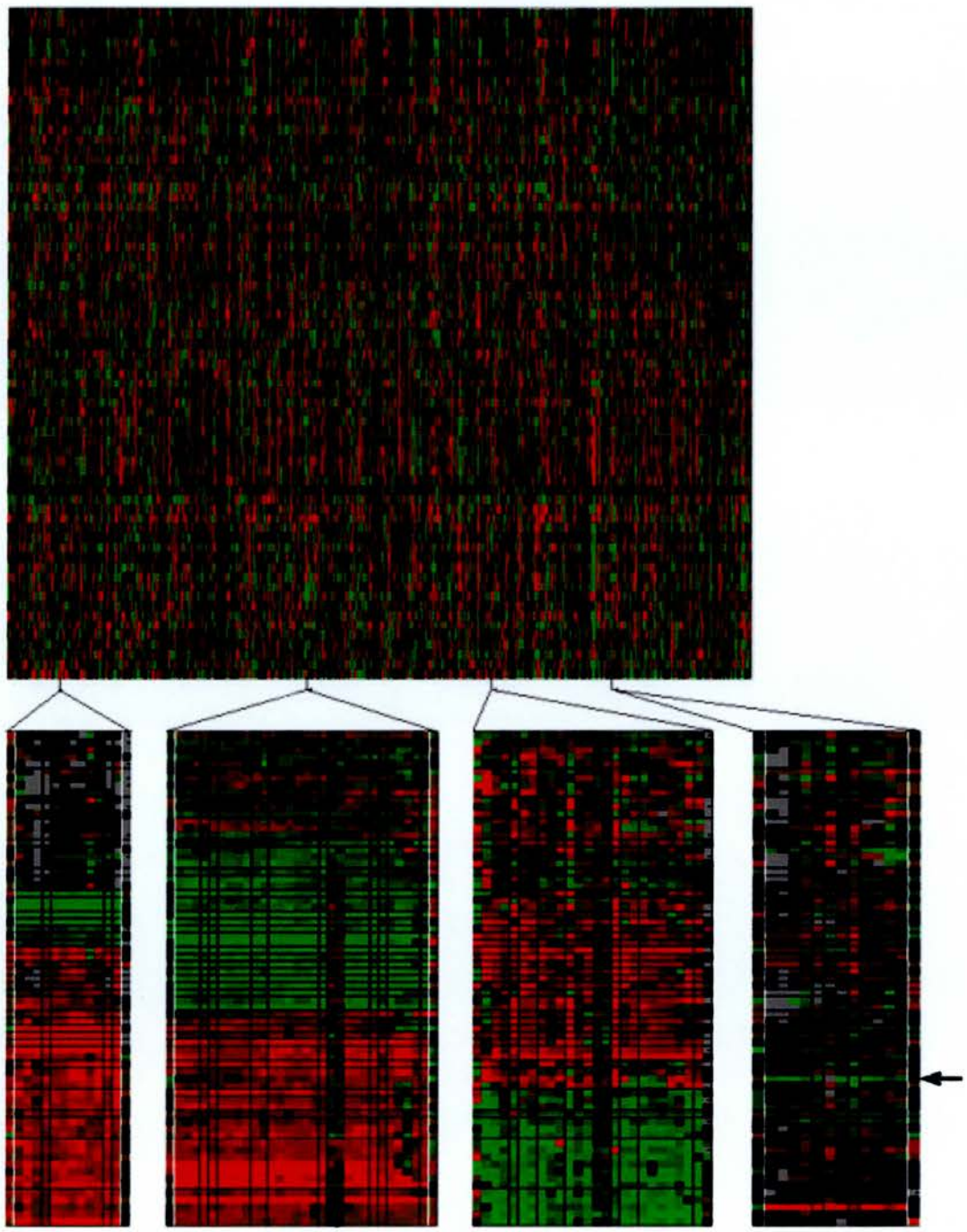


Figure 1.12 Microarray expression results
Microarrays can assess relative expression of two samples by using competitive hybridisation and fluorescent tags. The presented example is taken from an analysis of human chromosome 22 exons undertaken in order to validate them and infer transcriptional units. Each row represents an experiment using two samples to assess differential expression. The selected regions below the main results display each represent a gene, with each column corresponding to an exon. The region furthest to the right represents a novel transcript, with preferential expression in the testis compared to pooled sequences indicated by the arrow. Taken from Shoemaker et al., 2001.

Other forms of microarray continue to be developed, including microelectronic arrays from Nanogen (Edman et al., 1997; Sosnowski et al., 1997), which use electric fields to control hybridisation, Massive Parallel Signature Sequencing (MPSS) from Lynx Therapeutics (Brenner et al., 2000), which utilises fluorescent signatures to identify sequences, and fibre optic arrays from Gene Logic (Walt, 2000), which transmit fluorescence from labelled sequences to a detection system.

The major strength of array technology lies in the magnitude of genes that can be sampled at one time and microarrays allow analysis to be carried out on more than one sample at a time. However, this approach will not find novel genes and has low sensitivity, creating problems with reproducibility and necessitating a minimum two fold difference between samples for statistical robustness (Fryer et al., 2002).

1.3.9 *In silico* approaches

Informatic techniques have become an essential part of research activities in differential expression and the discovery of novel genes. This ranges from using the BLAST tool to elucidate a single gene corresponding to an experimental sequence, to complex programming that processes large amounts of data available from the public databases. The applications of these bioinformatic techniques include the elucidation and analysis of differentially expressed genes.

Various groups have designed computational methods for finding genes with tissue specific or preferential expression. Many have used a computer-based approach to analyse experimental data, whilst others have developed programs to mine public databases. The provision of similar tools via the internet have allowed this kind of analysis to become more accessible to a wide range of researchers. The two main programs available for this use are Digital Differential Display (DDD) at NCBI and a tool based on the gene indices at TIGR, which both facilitate identification of both

known and putative genes that show differential expression in the EST databases. These tools are described in more detail in a later chapter (section 3.1).

The TIGR method has been used successfully to identify testis specific genes (Olesen et al., 2001) in an attempt to elucidate novel genes that may be implicated in male infertility. This approach produced 131 novel testis specific sequences, with 14 being selected for RT-PCR analysis. Of these, four were specifically expressed, six were differentially expressed and four were ubiquitously expressed in the adult, with three of these being differentially expressed at the foetal stage.

A study aiming to identify genes important in cancer carried out a series of comparisons with normal tissues and breast, colon, lung, ovary, pancreas and prostate cancers using DDD (Scheurle et al., 2000). As well as producing known cancer-associated genes, 500 novel sequences were also derived from the analysis. Using RT-PCR, three of 12 known genes were shown to be specifically expressed, as were two of 20 novel genes. An additional example of the success of DDD is found in one of the recent studies showing the importance of Nanog as a factor defining ES cell identity, as the Nanog sequence was derived from one of the 20 highest EST clusters resulting from a DDD comparison of ES and somatic cells (Mitsui et al., 2003).

1.4 Progress in eye differential expression

Differential expression is a valid route by which genes important in disease can be found, either known or novel. This applies to ocular gene expression, as the majority of single gene eye disorders are characterised by a mutation in a preferentially expressed gene (Stohr et al., 2000). Many of the methods described earlier (section 1.3) have been applied to ocular sequences in order to assess the differential

expression of known genes and, in some cases, elucidate novel genes displaying preferential expression in eye tissue. This section describes results of investigations utilising the most common methods. The majority of these studies focus on retinal tissues, as this is the site of phototransduction and houses a variety of neuronal and non-neuronal cells with specific functions.

1.4.1 EST sequencing to identify novel differentially expressed genes

Many studies sequencing eye-derived ESTs do so from retinal libraries. Genes have been discovered in this manner, for example *FZD2-RS1* (frizzled homologue 2, related sequence 1, a homologue of *Drosophila frizzled*) (Wang et al., 1996) and *CLUL1* (clusterin-like 1) (Shimizu-Matsumoto et al., 1997), although these have as yet unknown function.

More restricted retinal regions are also used for library construction, such as fovea (Bernstein et al., 1996) and RPE (Paraoan et al., 2000; Buraczynska et al., 2002). In the RPE study by Paraoan et al., *CST3* (cystatin C), a protease inhibitor, was found to be differentially expressed by its representation in the constructed library and also by its representation in the databases, although this gene had no previous link with the RPE. This differential expression was subsequently confirmed by Northern analysis and RT-PCR. Another example of the discovery of preferential expression of a known gene was found by sequencing ESTs from bovine ciliary epithelium and using subsequent confirmational analysis to show that *GST-π* (glutathione S-transferase class-π), a member of a multifunctional enzyme family, was differentially expressed between different eye tissues and within the ciliary epithelium (Hernando et al., 1992). Other non-retinal EST sequencing efforts have produced novel genes, including *KRT12* (keratin 12), a structural protein differentially expressed in the corneal epithelium (Nishida et al., 1996), and an uncharacterised novel gene from human trabecular meshwork (Gonzalez et al., 2000a).

A major recent contributor of eye-derived ESTs has been the NEIBank project (Wistow, 2002). This project produced and sequenced thousands of ESTs from lens (Wistow et al., 2002c), iris (Wistow et al., 2002b), retina (Wistow et al., 2002d), and RPE/choroid (Wistow et al., 2002a) cDNA libraries derived from adult human eyes. These libraries were constructed to allow a resource of sequences representative of the mRNA population of these tissues and consequently were not normalised, although subsequent amplification and normalisation of the lens and iris libraries was carried out to reduce highly abundant clones before another round of sequencing was undertaken. Amongst the abundant ESTs found in this large-scale study were many novel and highly represented genes, in addition to novel transcripts of previously described genes. The sequences derived from these libraries will serve as a useful tool for future microarray studies.

EST sequencing has produced novel genes, however, differential expression has been deduced using further analysis on abundant sequences. Nevertheless, this technique has had success for the initial provision of sequences that can be tested.

1.4.2 Subtractive methods for assessing preferential expression and identifying novel differentially expressed genes

Subtraction has been employed on libraries derived from ocular tissues in order to produce tissue specific genes that may relate to tissue function. A retinal study, using Northern blotting to analyse expression of sequences from a subtracted library, found that only five of 23 novel sequences were not expressed in the retina and, of the rest, eight were expressed in the eye and less than four other tissues (Sinha et al., 2000). In an investigation of preferentially expressed sequences in the RPE, 45 sequences derived from a subtracted library were tested by Northern blot, with 29 preferentially expressed in the RPE (11 of these were specific and a further 11 were expressed in the RPE and retina only) and the remaining 16 showing retinal expression but no

expression in the RPE (Sharma et al., 2002). This neural retina expression was assumed to be caused by contamination of RPE tissue with neural retina prior to RNA synthesis. Another RPE study utilising a subtracted library identified two specific and one preferential sequence from seven novel sequences (Agarwal et al., 1995).

A differing subtraction technique synthesises cDNA from RNA and carries out a subtraction step before random priming from the remaining sequences (Rodriguez and Chader, 1992). This technique, known as solid-phase subtraction, was employed to find sequences preferentially expressed in the macula (Schoen et al., 1995). Of 100 sequenced clones, 63 were analysed by Northern analysis, 36% of which were not detected and exactly half of the remaining sequences were macula-enriched.

The SSH method has been applied in a study on two eye libraries, one retinal and one containing RPE and choroid (den Hollander et al., 1999). Of the 314 sequenced retinal cDNAs, 86 were novel and 21 of the 126 RPE/choroid sequences were novel. A semiquantitative RT-PCR method was used to assess the expression of 33 of these novel genes, yielding 11 specifically expressed sequences, 13 expressed significantly more in the expected tissues and nine expressed in the eye and in two or less other tissues.

These methods show that although there is variability in the efficiency of these techniques, they are successful at producing differentially expressed genes. Another study, carried out on a ciliary body library, demonstrates the applicability of this approach to questions about eye function (Escribano et al., 1995). From the 90 known genes corresponding to sequences derived from the human ciliary body subtracted library, a group of plasma proteins were identified. This expression was confirmed by Northern and *in situ* analysis and was a significant finding as it was previously thought that aqueous humour plasma proteins were synthesised outwith the eye.

1.4.3 A large scale *in situ* screen to identify differentially expressed genes

A high throughput *in situ* hybridisation procedure was utilised in an effort to gain molecular markers for development of the anterior eye (Thut et al., 2001). Probes were made from 1,035 ESTs, with each tested at six stages of mouse development. From the results, 62 probes displayed distinct expression in the RPE, outer neural retina, inner neural retina or the presumptive iris and ciliary body. These developmentally expressed sequences were used as markers for inductive signals in the anterior eye that may lead to lens induction and were tested in lens culture studies by *in situ* hybridisation. This study represents a large-scale project that aims to answer a specific developmental question and has succeeded in producing candidates for factors involved in lens development.

1.4.4 Using differential display to assess preferential expression and identify novel differentially expressed genes

The DD method has been used to identify differentially expressing eye genes, particularly in the retina and lens. This method was used in a study aimed at finding genes that were upregulated following severance of the optic nerve in rat, as they hypothesised that the RGC cell death following this process is caused by the induction of apoptotic genes (Levin and Geszvain, 1998). Using DD, upregulation of ceruloplasmin, an antioxidant, was found in the retina after surgical optic crush. This was confirmed by Northern blotting and RT-PCR showed differential expression in normal retinas, with expression in the liver and, to a lesser extent, in the retina. Immunofluorescence was used to show expression in the INL and RGCs.

Ceruloplasmin therefore represents a candidate for inducing RGC cell death in the retina.

A study on chick retina and choroid utilised DD to assess differential expression in chick eyes that had been subjected to induced refractive error (Feldkaemper et al., 2000). This was done by using lenses and fogged goggles to change the focus of the lens and had previously been shown to induce the relevant change in cornea thickness. After 6-24 hr of treatment, 12 retinal and five choroidal genes were seen to be differentially expressed. Two genes, a cytochrome subunit and proglucagon, were shown to be upregulated and these were confirmed by Northern blotting.

A study investigating differential expression between cultured and fresh RPE cells highlights the alterations of gene expression in culture, but was used in the investigation as an indicator of genes that may be important in a disease characterised by the loss of epithelial nature of RPE cells (PVR, proliferative vitreoretinopathy), as occurs in cultured tissue. A candidate gene, *MAP1B* (microtubule-associated protein 1B), linked with roles in cell or subcellular structure determination, was found using this method, as it was upregulated in culture compared to fresh RPE and protein localisation showed it to be expressed in PVR, but not in disease-free, RPE.

There has been evidence of differences between sexes in humans and in animal models with relation to developing age-related cataract. DD has been used to assess this difference in culture with male and female lenses and also with TGF-β (transforming growth factor, beta), known to induce greater susceptibility to cataract in male rats and females with removed ovaries (Sun et al., 2000). Two known genes were highlighted in this study. *Crygb* (gamma crystallin B), a structural protein, was found to be upregulated in males lenses and proteasome Z, a subunit of the small peptide hydrolysing proteasome, was upregulated in TGF-β induced male lenses but downregulated in ovariectomised female lenses and TGF-β female lenses compared to normal lenses. These two genes found by DD represent candidates for sex-related differences in the propensity for age-related cataract development. Another cataract

study used DD to compare age-related cataract-affected lenses and with cataract-free lenses and found three genes upregulated in the cataract, including *METII* (metallothionein IIa), a detoxification protein induced by oxidative stress (Kantorow et al., 1998). One of the 12 genes found to be downregulated in cataract-affected lenses was *P2A-RS* (protein phosphatase 2A regulatory subunit), a mitotic suppressor.

In a study to find genes preferentially expressed in undifferentiated anterior lens epithelium cells, DD was used to find differentially expressed genes between untreated anterior cells in culture compared to those treated with the differentiation factor bFGF (Cai et al., 1999). A novel gene was expressed preferentially in undifferentiated cells and subsequent Northern analysis confirmed its presence in anterior cells but only trace amounts in differentiated LFCs.

DD has allowed the direct comparison of two or more samples, showing upregulation as well as downregulation, therefore presenting an ideal opportunity to study diseases affecting the eye. These studies have mainly been focused on assessing differential expression of characterised genes, but novel genes have also been found in this way.

1.4.5 Using SAGE to assess preferential expression

SAGE studies have been carried out on retinal and corneal tissues. A large SAGE study constructed libraries for adult hypothalamus, retina at two day intervals between E12.5 and P2.5, adult retina, adult ONL and postnatal retinas from normal and homozygous mice for *Crx*, a homeobox gene expressed in developing photoreceptors (Blackshaw et al., 2001). The SAGE libraries were compared to deduce genes that may be important in photoreceptor development and function. Of the 264 preferentially expressed sequences corresponding to novel genes, 86 mapped to 37 uncloned retinal diseases. Another retinal SAGE study compared libraries

constructed from human peripheral retina, macula and RPE (Sharon et al., 2002). Comparisons of these with non-ocular libraries from the public domain gave 89 genes that are preferentially expressed in the retina, 53 of which are RPE specific. In comparisons between their SAGE libraries, genes encoding axon structural proteins were upregulated in macula compared to the periphery. Tissues that were used for SAGE libraries were taken from one 44 and one 88 year old, with genes expressing iron metabolism and oxidative damage protection proteins upregulated in the latter.

The most recent reported ocular SAGE libraries reported were constructed in order to assess differential expression between the normal human corneal endothelium and that affected by Fuch's dystrophy, a condition characterised by cell irregularity and areas of thickened Descemet's membrane (Gottsch et al., 2003). From a total of 19,136 sequenced tags, 18 sequences were deduced to be upregulated in Fuch's dystrophy corneas and 36 were downregulated. Of these downregulated sequences, seven showed no matches in the databases, three matched ESTs and the remaining 26 known downregulated genes have known antioxidant and toxic stress management properties.

SAGE libraries have been used to show differential expression between libraries in all three studies, but an important function of these libraries is for future mining and for the production of novel sequences that may be candidates for ocular disease.

1.4.6 Using arrays to assess preferential expression

A large range of array types have been utilised in ocular research. One study used this approach in order to gain insight into genes expressed in the human cornea by using a corneal target on an array consisting of 5,600 human genes (Jun et al., 2001). Differential hybridisation and macroarrays have also allowed the elucidation of novel genes expressed in human photoreceptors (Swanson et al., 1997) and in chick

(Hackam et al., 2003) and canine (Lin and Sargan, 2001) retina. However, microarrays are designed to deduce differentially expressing sequences and therefore this is their major use.

Microarray studies of retinal disorders have yielded candidate genes for disorders and genes that may be involved in a disease phenotype. Retinal examples of this in humans are *SFRP2* (secreted frizzled-related protein-2), previously postulated to be influential in retinal cell polarity and found to be upregulated in human retinitis pigmentosa (a degenerative retinal disease) retinas (Jones et al., 2000a), a range of genes upregulated in rats with induced diabetes suggesting an inflammatory component to the response (Joussen et al., 2001) and upregulation of genes in ageing (62 to 74yr) retina compared to young (13 to 14yr) retina (Yoshida et al., 2002). In mouse, upregulation of a set of known genes was found in the *rd/rd* mouse (with a mutation in the homologous gene to that involved in retinitis pigmentosa in human) before its characteristic morphological retinal degeneration (Jones et al., 2000b) and light-induced apoptosis in a mouse strain containing a sensitising mutation correlated with downregulation of many photoreceptor specific genes and an increase in stress response gene expression before morphological changes (Choi et al., 2001).

Corneal studies have also produced candidates for genes involved in human diseases such as bullous keratopathy, a corneal disorder that commonly occurs after cataract removal and leads to corneal opacity (Spirin et al., 1999), and keratoconus, a condition that causes myopia and astigmatism and is characterised by corneal thickening and an irregular epithelium (Nielsen et al., 2003). Glaucoma, associated with an increased intraocular pressure, has been investigated using arrays to show higher expression of a set of known genes (Gonzalez et al., 2000b).

As well as characterising ocular phenotypes further with gene expression analysis, mouse strains with defects in eye genes can be used to investigate transcriptional networks and developmentally expressed genes. Comparisons between mouse $Crx^{-/-}$ and $Crx^{+/+}$ retinas (Livesey et al., 2000) and lenses from $Pax6^{-/-}$ and $Pax6^{+/+}$ mice

(Chauhan et al., 2002) have identified downstream targets of these genes in the relevant tissues.

1.4.7 *In silico* approaches to assess preferential expression and identify novel differentially expressed genes

The EST databases have been mined successfully in order to elucidate sequences preferentially expressed in the eye or in specific eye tissues. One of these efforts was based on the entire dbEST database, using an in house *in silico* subtraction technique to derive 925 sequences preferentially expressed in the retina and providing candidates for 49 of the 51 uncloned retinal diseases (Katsanis et al., 2002).

Two studies have utilised the TIGR database, with the first taking ESTs from retina specific clusters and mapping 55 of 83 novel sequences to 14 candidate regions (Malone et al., 1999). The second TIGR study focused on retinal and pineal gland libraries, with the justification that pinealocytes are developmentally related to photoreceptors and some genes, such as *CRX*, are expressed only in the retina and pineal gland (Sohocki et al., 1999). On this basis, THCs with more than 10% EST representation in pineal gland libraries were picked from TIGR and only those with additional retinal transcripts but no representation in other tissues were extracted. Of these 45 clusters, 22 were deemed to represent novel genes expressed only in the retina and pineal gland.

Another two studies utilised UniGene rather than TIGR for this approach, although neither used the DDD function but instead used techniques they developed. The first based preferential expression on the ratio of retinal EST representation in a given gene cluster compared to the total retinal EST population (Bortoluzzi et al., 2000), a similar approach to that used by the UniGene utility, DDD (section 3.1). Using this method, 536 retina specific clusters and 443 clusters preferentially expressed in the

retina were hypothesised as representing the retinal transcriptome, as allowed by available sequences in UniGene. The second study utilising UniGene extracted the 1,241 clusters with greater than 30% retinal EST composition, including 673 specific clusters, and after *in vitro* analysis of 180, found 39 to be retina specific and another 30 preferentially expressed (Stohr et al., 2000). Additional investigation of a cluster characterised it as a novel gene displaying alternative splice forms.

Computational approaches present a relatively new method for assessing preferential and specific expression in the eye and have led to the discovery of novel genes. However, to date, there has been little characterisation of novel genes predicted to be preferentially expressed in the eye by *in silico* methods. Therefore the potential for gene discovery by this method has not yet been fully realised and presents the opportunity for further research in this area.

1.5 Identification and expression of genes implicated in eye disease

This section describes the isolation and characteristics of a number of genes that underlie eye diseases. They each differ in their expression pattern, and are representative of types of genes that may be, or may not be, identified by Digital Differential Display.

1.5.1 *Rho*

Rho encodes rhodopsin, a protein expressed only in the retina, which was long known to be involved in visual function and was an obvious candidate for a gene underlying visual defects.

As discussed earlier (section 1.1.5), rhodopsin is an essential component of phototransduction, activated by light to initiate the excitation phase. It was first isolated as the pigment visual purple in 1878 (Garriga and Manyosa, 2002), which was later named rhodopsin and shown to be composed of a protein, opsin, and a chromophore. Opsin is a 40kDa apoprotein, encoded by the *RHO* gene, that is bound to 11-*cis* retinal (Elliott et al., 1990). *RHO* was sequenced in 1984 (Nathans et al., 1986) with the mouse gene isolated in 1988 (Baehr et al., 1988).

Rhodopsin protein is found exclusively in rod photoreceptors of human and mouse, with the *Rho* gene being specifically expressed in the eye. RHO is a member of the G-protein coupled receptor (GPCR) superfamily of transmembrane proteins that bind and activate a G-protein. Specifically, rhodopsin binds and activates transducin in the rod OS, where it is associated with the flattened membranes, known as discs, within the OS. As well as being a model of GPCRs, rhodopsin is also associated with retinal disease.

There are two human retinal disorders associated with mutations in *RHO*: retinitis pigmentosa (RP) and congenital stationary night blindness (CSNB). RP is characterised by progressive degeneration of the retina, beginning with night-blindness and a loss of peripheral vision and eventually loss of central vision. Both rod and cone cells are lost through degeneration, probably through apoptosis (Jomary et al., 2001, (Yang et al., 2001). Altogether, over 100 mutations in *RHO* have been described and are associated with approximately 50% of all inherited RP in humans (Chapple et al., 2001). The most common form of RP associated with rhodopsin is autosomal dominant RP (ADRP), with 38 mutations reported in OMIM. The

majority of these dominant mutations in *RHO* are thought to cause RP by causing incorrect folding of opsin or through non-localisation to the OS discs with consequent degeneration (Berson, 1996). Three null mutations described in OMIM are associated with autosomal recessive RP. The difference between the dominant and recessive forms of RP caused by *RHO* mutations is likely to be due to the type of mutation. Mutations causing dominant RP produce abnormal RHO protein, which is thought to lead to degeneration of photoreceptors. However, null mutations will not produce RHO protein, therefore those heterozygous for the mutation may have a reduced amount of rhodopsin but degeneration of rods may not be induced. Instead, human carriers of putative null mutations have virtually normal vision before the onset of significant photoreceptor death (Jacobson et al., 1996; Kajiwara et al., 1993).

Only three dominant mutations in *RHO* have been associated with CSNB (Garriga and Manyosa, 2002). These mutations have been associated with the disruption of maintaining inactivity, therefore these mutations may cause constitutive activation of rhodopsin (Garriga and Manyosa, 2002). It has been proposed that this inappropriate stimulation of rod cells leads to their desensitisation (Jin et al., 2003).

Two mouse models with disrupted endogenous *Rho* have been generated. Both homozygous knockouts displayed failure of rod OS formation, with complete degeneration of photoreceptors within a few months after birth (Humphries et al., 1997; Lem et al., 1999). Heterozygotes in one study developed OS, but with half of the normal amount of rhodopsin present, with retinal degeneration occurring over a much longer period of time (Lem et al., 1999). The other study produced heterozygous mice that retained most of their photoreceptors, although the inner segments (IS) and OS were structurally abnormal (Humphries et al., 1997).

As *RHO* is expressed specifically in the rod photoreceptors, diseases associated with a mutation in this gene result specifically in the retinal diseases RP and CSNB.

1.5.2 *Pax6*

Pax6 is an example of a gene that is involved in eye disease, but which is expressed in both the developing eye and in the brain.

PAX6 (paired box gene 6) was first identified as a candidate for aniridia in humans by positional cloning (Ton et al., 1991). This type of aniridia is characterised by a range of ocular abnormalities that can include hypoplasia of the iris, fovea and optic nerve, as well as cataracts and nystagmus, an oscillation of the eyeballs. Mouse *Pax6* was subsequently identified as the gene containing the *small eye* mutation in mice (Hill et al., 1991). This is a semidominant mutation, with heterozygotes having smaller eyes than wild type and often displaying an optic fissure. Homozygotes have no visible eyes and undeveloped nasal cavities and therefore die soon after birth.

Pax6 is expressed embryonically in a restricted set of tissues other than the eye; nasal structures, brain and neural tube (Walther and Gruss, 1991; Grindley et al., 1995). In the adult, it is expressed in the eye, cerebellum and pancreas (Ton et al., 1992; Turque et al., 1994; Stoykova and Gruss, 1994). Its sequence is highly conserved across invertebrates and vertebrates and contains a homeodomain as well as a paired box domain, separated by a linker portion and followed by a PST domain (proline, serine and threonine rich) that has functions in transcriptional activation (Glaser et al., 1994). Within the eye, *Pax6* is implicated in eye development through its proposed regulation of several developmentally important genes. These include transcription factors such as *Six3*, *Prox1* and basic helix-loop-helix transcription factors that are important in cell determination and differentiation, keratins and crystallins. Through these interactions, *Pax6* is thought to be involved in cell cycle regulation, cell determination and differentiation and apoptosis (Simpson and Price, 2002).

Many *PAX6* mutations have been described, with the majority causing aniridia, a disease with a population frequency of 1 in 60,000 to 1 in 100,000 (Prosser and van,

V, 1998). Although the majority of anridia cases have been shown to be caused by a *PAX6* mutation, there are also anirida patients with WAGR (Wilms tumour, Aniridia, Genitourinary abnormalities, mental Retardation), a syndrome caused by deletions of 11p13 including *PAX6* (Fantes et al., 1992). The paired and homeodomains of *PAX6* are the source of a large proportion mutations in *PAX6* when considering their size (Prosser and van, V, 1998). The vast majority of phenotypes resulting from *PAX6* mutations appear to be caused by a loss of function, and cause predisposition to aniridia. However, other eye abnormalities are seen, particularly in patients displaying milder phenotypes, and may be caused by altered specificity of the protein.

In contrast, there are fewer mutations in the PST domain, which is thought to be at least partly due to lower detection caused by the milder effects on *PAX6* activity (Prosser and van, V, 1998). An example of this is a phenotype caused by compound heterozygosity, with members of a family displaying congenital cataracts and late-onset corneal dystrophy whilst others suffered from aniridia (Glaser et al., 1994). This was associated with two distinct point mutations in the PST domain. There are non-coding regions of the PAX6 gene that have conservation across species but, to date, no mutations have been found in these regions.

There are two alternative transcripts produced by the *PAX6* gene by either splicing in or leaving out exon 5a (Azuma et al., 1999). These two forms differ in the paired domain region, which consists of two DNA binding subdomains known as the N terminal subdomain (NTS) and the C terminal subdomain (CTS). With exon 5a included, there are 14 additional residues in the NTS, removing its binding activity and allowing the binding activity of the CTS. This introduces an element of target specificity by the PAX6 proteins. This extra exon is specific to vertebrates and a mouse null for *Pax6* exon 5a displayed iris hypoplasia as well as defects in the cornea, lens and retina (Singh et al., 2002). This difference from the anopthalmic phenotype, coupled with CNS abnormalities and subsequent lethality, suggests a distinct function for this isoform in the development and function of a specific subset of ocular structures.

There are many mouse models of eye disease caused by *Pax6* mutations such as the original *small eye* mutation mentioned earlier. Some of these mutations have been induced, including small eye Neuherburg (Hill et al., 1991), while others are spontaneous, such as Dickie's small eye (Theiler et al., 1978). These are all semidominant mutations and produce broadly similar phenotypes but with variable severity.

1.5.3 Crystallins

Crystallins were first identified as major components of the lens. They are soluble proteins that comprise two types of gene product; some crystallins are eye specific in their expression whilst others are widely expressed but are more abundant in the eye.

As previously mentioned, lens thickness increases throughout life, with cells proliferating continuously from the anterior lens cells and accumulating at the periphery, leaving the oldest cells in the centre of the lens (section 1.1.3). Lens transparency is required for clear vision to allow the passage of light without scattering and this state must be maintained in the oldest cells as well as in the new. This is achieved by the destruction of organelles within lens cells and also the highly regular structure formed by lens proteins, a large proportion of which are crystallins.

The crystallins are soluble proteins that form the major structural components of the lens and are important in lens transparency. This function is often recruited and a phenomenon known as gene sharing, when crystallins carry out different functions in other tissues, is common in the lens (Piatigorsky et al., 1988; Piatigorsky and Wistow, 1989). In these cases, expression levels in other tissues are much lower than they are in the lens. The most common functions for crystallins in non-lens tissues are as chaperone proteins or metabolic enzymes (Wistow and Piatigorsky, 1988; de Jong et al., 1989).

Not all crystallins have had expression detected in other tissues. Two subunits of alpha crystallin, alpha A crystallin, encoded by *CRYAA*, and alpha B crystallin, encoded by *CRYAB*, are examples of crystallins expressed in a restricted set of tissues and more widely expressing crystallins. Alpha B crystallin has the more widespread expression of the two, and is found in many tissues including heart, kidney, skeletal muscle, retina, lung and brain, as well as lens (Bhat and Nagineni, 1989; Dubin et al., 1989).

Alpha crystallins appear to interact with lens cytoskeletal proteins phakinin and filensin (Muchowski et al., 1999). The exact role of alpha crystallins is unknown, however their function as chaperone proteins in other tissues and their required presence for correct folding of cytoskeletal proteins has led to hypothesised function in cytoskeletal assembly (Carter et al., 1995; Hess et al., 1998). It is thought that alpha crystallin selectively binds unfolded or denatured proteins, particularly prevalent in the ageing central fibres in the lens, to suppress non-specific aggregation which may follow and disrupt mammalian lens transparency (Wang and Spector, 1994; Boyle and Takemoto, 1994; Rao et al., 1995; Carver et al., 1996).

Cataract describes the opacification of the lens and reduces visual acuity and contrast sensitivity. There are both age-related and congenital forms of cataract. Crystallin genes are often associated with cataracts in the human population. Mutations in both alpha A and alpha B crystallin have been discovered in humans with cataract phenotypes.

Both dominant and recessive forms of cataract have been associated with mutations in the *CRYAA* gene. One dominant form has a point mutation within a highly conserved residue (Litt et al., 1998). Using transfection studies with lens epithelial cells, the protein was found to localise to the nucleus, which is abnormal, and these cells were not protected from apoptosis as control cells were (Mackay et al., 2003). The recessive mutation in *CRYAA*, identified in an inbred Jewish Persian family, was found to be a nonsense mutation (Pras et al., 2000).

The phenotypes associated with *CRYAB* mutations are more diverse than those with mutations in *CRYAA*. The first mutation discovered in *CRYAB* was associated with dominant desmin-related myopathy with cataract (Vicart et al., 1998). Affected members of the family in which this mutation was found had limb muscle weakness, cardiomyopathy and cataract. Desmin is a type III intermediate filament protein and muscle biopsies of family members with this disease displayed aggregates of desmin. Transfection of muscle cells with this mutant *CRYAB* is associated with aggregation of desmin and alpha B crystallin. Another mutation in *CRYAB* was shown to cause autosomal dominant posterior pole cataract in a family, with the mutation causing a frameshift and altering the final 35 residues (Berry et al., 2001). The authors speculated that this may occur through an increased tendency for the mutant protein to aggregate or from a loss of chaperone activity.

Alpha A crystallin knockout mice develop cataract (Brady et al., 1997). Further investigation of the lenses of these mice indicate the formation of aggregates predominantly containing alpha B crystallin but also composed of gamma crystallins. It is hypothesised that the decrease of soluble gamma crystallin in alpha A crystallin knockout mice may contribute to the cataract phenotype observed.

1.5.4 *Rpgr*

Mutations in the RPGR gene cause an eye disease, although the gene is expressed ubiquitously.

RPGR (retinitis pigmentosa GTPase regulator) was first identified as harbouring loss of function mutations in patients with RP3 (retinitis pigmentosa 3), an X-linked disorder (Meindl et al., 1996). In this study, two intragenic deletions, two nonsense mutations and three missense mutations were found to be associated with RP3.

The RPGR protein contains a series of tandem repeats, with eight potential glycosylation sites and a consensus isoprenylation site at the C terminus, the latter thought to be a potential region for membrane anchorage (Meindl et al., 1996).

Although RPGR is ubiquitously expressed, there are many alternative splice forms, some of which are tissue specific, including an eye specific transcript containing exon15A that introduces a stop codon and truncates the resulting protein, including the isoprenylation site (Kirschner et al., 1999). More exons have been discovered since, including a conserved 3' terminal exon known as ORF15 (Vervoort et al., 2000) that has since been shown to be mutated in 80% of X-linked retinitis pigmentosa (XLRP) patients tested in Britain (Vervoort and Wright, 2002). This exon is preferentially expressed in mouse and bovine retina and encodes 567 additional residues, including a glutamic acid rich repetitive domain. Due to the high proportion of XLRP cases that can be attributed to this region, it is considered to be a mutation hotspot.

In a study of 3 unrelated Japanese families with a mutation in ORF15 predicted to cause truncation of the RPGR protein, affected males displayed RP whilst female carriers were more varied in the severity of the phenotype (Yokoyama et al., 2001). Some of these females had RP and most displayed myopia and astigmatism.

Further evidence of a wider range of phenotypes associated with *RPGR* mutations was provided in a study describing males in a family where the underlying point mutation was found in an intron and was thought to create a donor splice site (Ayyagari et al., 2002). Ten males in this family displayed macular atrophy resulting in a progressive loss of visual acuity but with only minimal peripheral visual impairment, with cone and rod responses appearing normal in some of these family members. However, one male showed more extensive macular degeneration and peripheral loss of the RPE.

In a study that found 80 mutations in *RPGR*, 41 were novel (Sharon et al., 2003). Of the 187 unrelated males with XLRP involved in the study, 66% had a mutation in

ORF15. Further investigation revealed that the final dark-adapted vision became lower, thereby being closer to normal, as the length of wild type ORF15 sequence increased. Additionally, it was shown that the severity of the disease increased as the abnormal sequence length following mutation in ORF15 increased. The ORF15 mutations were associated with milder forms of XLRP compared to mutations in exons 1-14.

In RPGR knockout mice, cone opsins were localised incorrectly to the cell body and synapses of the cone and rods contained a reduced level of rhodopsin (Hong et al., 2000). Degeneration of rods and cones was observed. RPGR had previously been localised to the rod OS in cow and human (Roepman et al., 2000) and also to cone OS in humans (Mavlyutov et al., 2002). In the knockout study, RPGR was found to normally localise specifically to the connecting cilia of rods and cones. For this reason, it was hypothesised that RPGR is involved in maintaining the polarised distribution of protein across the connecting cilium, either by directional transport or by restricting redistribution.

1.6 Aims

This thesis describes a method to elucidate novel genes that may be influential in the manifestation of eye diseases. As specific or preferential expression of a gene in a particular tissue or organ is often indicative of its importance in function or development, this basis was used to discover such genes. This approach may allow the discovery of specifically or preferentially expressed genes, such as *Rho*, *Pax6* and *Cryaa*, but will not be expected to highlight widely or ubiquitously expressed genes, such as *Cryab* or *Rpgr*.

An *in silico* approach was employed to produce a set of potential candidate genes for disease relevance. Subsequently, further bioinformatic and *in situ* analyses were utilised to further investigate these possibilities.

CHAPTER 2

Materials and Methods

2.1 Bioinformatics

The following databases have been used in producing or verifying the data shown in this thesis:

Databases	URL
UniGene	http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unigene
Ensembl	http://www.ensembl.org
MGI	http://www.informatics.jax.org
OMIM	http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM
Swiss-Prot and TrEMBL	http://www.ebi.ac.uk/swissprot/
SMART	http://smart.embl-heidelberg.de/

The following tools have been used in producing or analysing data shown in this thesis:

Bioinformatic tool	Reference/URL
Primer	http://www-genome.wi.mit.edu/cgi-bin/primer/primer3.cgi/
BLAST	Altschul et al., 1990
ClustalW	Thompson et al., 1994
Mega	Kumar et al., 2001
VISTA	Mayor et al., 2000

2.2 Manipulation of nucleic acids

2.2.1 Extraction of RNA from tissue

The RNA was extracted from tissue by using RNA-Bee (Tel-Test). The procedure was carried out according to the provided protocol with the following specifics:

- Following chloroform addition, samples were stored on ice for 5 min.
- Following isopropanol addition, samples were left at room temperature for 10 min.
- Centrifugation steps following isopropanol and 75% ethanol additions to the sample were carried out at 4°C.
- RNA was resuspended in 25 µl DEPC dH_2O.
- The amended protocol for small amounts of tissue was used for extracting RNA from E14.5, E16.5, neonate and adult eyes and ovary.

2.2.2 DNase treatment of RNA

RNA samples were DNase treated using DNA-free (Ambion) according to protocol with the following specifics:

- RNA samples were incubated at 95°C for 3 min before placing on ice briefly and beginning the DNase treatment.
- 2.5 µl of 10 × DNase I buffer was added to the RNA sample.

- 5 µl of DNase Inactivation Reagent was added to the sample following incubation at 37°C.
- RNA was removed from the DNase Inactivation Reagent immediately after centrifugation.

2.2.3 Determining the concentration of RNA

RNA samples were quantified by using the Genequantpro spectrophotometer with samples diluted 1 in 100.

2.2.4 Primer design

Primers were designed using the Primer3 programme (see section 2.1) using default settings, with the exception of stipulating an amplification length of 300bp to 350bp. The following forward and reverse primer sequences were synthesised (MWG Biotech) and used in the course of producing results seen in this thesis:

Cluster identifier	Forward primer sequence	Reverse primer sequence
Mm.103712	ACAGTTCCACAGTTGGGAGG	GCAGACACCTCTGTTAGCCC
Mm.121647	AACATTTTTGTGAGGCAGGG	CCAGGATGCTCAAAGGAAAA
Mm.134224	TGGCTTCTGCCCATATAACC	CCTAAGCTCCAGAGACCCCT
Mm.134516	ACCAGCAACCTGAAATGGAC	ATCCACACAGCCCTGAAAAG
Mm.137178	TTAAAGAGGTCGCCGAAGTG	AGAAACTGCTTGGTGCCAGT
Mm.150838	TGTCACCTCTTCTGGGGAAC	TGTCGGGGGATCACTTCTAC
Mm.151918	GGCTGGTGGAATTCAAAGAA	TTCCTTCCATTTTTGGCAGT
Mm.159861	TCCCCACTAGAATGACCTTG	CGACTAACGCTAGGCAAAGG
Mm.195923	CCCACTCCACTTCTCACAGG	CTTCCAAATGATGAAAGGGC
Mm.200734	ACAAACAGCGATTTGGCTTG	AATCTGGCTTCCCTGTCCTC
Mm.213208	TAAGACAGGAGACCCCATCG	AGGGCCAGCCTTTACTTCAT
Mm.21485	ACCTCCACTCTTCCCTCCAT	CATCCCTGAGAAGAGGCTTG
Mm.22913	AATGCAGCTTCATGCCTACC	GCACAGAGTCAGGTCAGCAG
Mm.23434	GACCAAGGAGTCAAGGGACA	AAATGTGCCCCTGACAGAAG
Mm.2965	TTAGAAACAGTGGAACCGGC	TCACATCAGCTTTAATGGGAA
Mm.29846	CAAACAAAGGTTCCATGCCT	TCAGTATTCACGGAAGCACG
Mm.3124	CAGTCGTAAAGCAGCACAGC	CAGACTGGAAAGCCAGGAAG
Mm.34901	TCAATGAGGGCTCCAAAATC	AGCAGGAACGTTGTGGTAGG
Mm.35511	CGAGCCAGTTTCTGAAGGTC	TCAGGCCAGATGCAGAGTC
Mm.37819	GCTTGCAGGTTATTTCTGCC	AGCTCTTGGGTCAGTGCTGT
Mm.38347	AATACCTGGTGGAGAATGCG	CTTCAGTACCTGGCCCCTCT
Mm.44242	CCTCTCCTGTCATGCTGTGA	CTTGGTGGGACAGGATATGG
Mm.44292	GGGCTCTCAGCCCTAAAACT	CCAAATACCCAATTACCCCC
Mm.56565	ACCTGTAATCTCCGCACTCC	TTTAATCACGCAGCTCAGGA
Mm.59151	CTGCTACCATTTCCATGCCT	AACAATCCACCAGGTTCCAG
Mm.62589	ATCACCCCAAGGTTGATCTG	TGAGTTCACAGCAAACCCAG
Mm.78118	GCACCTACCAAACTGGCATT	TGAGATATGTAACAATTTCTTTGGG
Mm.82341	AGCCTGTATTGAGGACCCTG	CCTTTATATACCTCCATACCTGGG
Mm.87130	TCAGTAGGACCGAAGCATCA	ATAAGGATCGTGGAAGCCAA
Mm.95707	AACTACTGCCAGTGCCTGCT	CCACCCAACCTAACCCAGT
Mm.95741	ACCCTCACACTCAGGTGGAA	GAAGGCCAATTTGCCACA
Gapd	CCCACTAACATCAAATGGGG	ATCCACAGTCTTCTGGGTGG

2.2.5 Polymerase chain reaction

PCR reaction mixtures were made up for multiple reactions and then aliquoted, with the following amounts of components per reaction to produce a 1.5 mM Mg^{2+} concentration:

Component	Volume (μl)
Taq polymerase 5 U/μl, Applied Biosystems	0.2
10 mM dNTPs 40 μl each of 25 μmol dA-, dC-, dG- and dTTP, (ABgene) plus 240 μl dH_2O	0.5
100 ng/μl forward primer 10 μl of 1mg/ml synthesised primers in dH_2O plus 90 μl dH_2O	0.5
100 ng/μl reverse primer 10 μl of 1mg/ml synthesised primers in dH_2O plus 90 μl dH_2O	0.5
$MgCl_2$ 25 mM, Applied Biosystems	1.5
10 × buffer Applied Biosystems	2.5
dH_2O	18.3

Added 1 μl of template DNA, or dH_2O for negative control, to each aliquot of PCR mix to give a final reaction volume of 25 μl. The following programme was used on a Peltier Thermal Cycler DNA Engine DYAD unless otherwise stated:

Temperature (°C)	Time (min)	
94	5.0	
94	1.0	} 35 cycles
55	1.0	
72	1.5	
72	10.0	

2.2.6 Colony PCR

A picked colony was introduced to 20 μl dH_2O, which was then incubated at 100°C for 5 min on a Peltier Thermal Cycler DNA Engine DYAD. This colony DNA was used as the template for colony PCRs.

2.2.7 Reverse transcription polymerase chain reaction

RT-PCR was carried out in two steps. The following component were mixed on ice to set up the 20 µl reverse transcription reaction, with the RNA and DEPC dH_2O volumes totalling 10.5 µl:

Component	Volume (µl)
5 × AMV-RT buffer Roche	4.0
10 mM dNTPs 40 µl each of 25 µmol dA-, dC-, dG- and dTTP plus 240 µl DEPC dH_2O	2.0
10 × hexanucleotide mix Roche	2.0
RNase inhibitor 40 U/µl, Roche	0.5
Reverse transcriptase, AMV 25 U/µl, Roche	1.0
1 µg/µl RNA	X_1
DEPC dH_2O	X_2

For each of these reactions, another reaction was set up in which the reverse transcriptase was replaced with an equal volume of DEPC dH_2O to allow for a negative control at the PCR stage. These reactions were all incubated at 42°C for 1 hr followed by 8 min at 75°C to stop the reaction. PCR reaction mixtures were then made up for multiple reactions and then aliquoted, with the following amounts of components per reaction to produce a 1.2 mM Mg^{2+} concentration:

Component	Volume (µl)
Taq polymerase	0.10
10 mM dNTPs	0.50
100 ng/µl forward primer	0.50
100 ng/µl reverse primer	0.50
$MgCl_2$	0.95
10 × buffer	2.00
dH_2O	13.45

Each aliquot of PCR mix was added to by 2 µl of reverse transcription reactions as template DNA, or dH_2O for negative control, to give a final reaction volume of 20 µl. The following programme on a Peltier Thermal Cycler DNA Engine DYAD was used:

Temperature (°C)	Time (min)	
94	5.0	
94	1.0	} 30 cycles
60	1.0	
72	1.5	
72	10.0	

2.2.8 Agarose gel electrophoresis and determining the concentration of DNA

Agarose gels were prepared using the appropriate proportion of agarose ('Hi-Pure' Low EEO Agarose, Biogene) in 1 × TBE for 0.8%, 1.0% or 1.5% gels as required, with the following recipe per litre of 20 × TBE (by MRC Human Genetics Unit technical staff):

Component	Amount
TRIS base	216 g
Boric acid	110 g
0.5 M EDTA	80 ml

The equivalent of 3 µl ethidium bromide (BDH Laboratory Supplies) for every 50 ml of gel volume was also added. 0.5 µl orange/blue 6 × loading dye (Promega) was added to each 2.5 µl of nucleic acid to be run, to a maximum of 2 µl for a 10 µl sample. The markers that were run alongside samples to determine their size were 100bp DNA Ladder (Promega), 1kb DNA Ladder (Invitrogen Life Technologies) or bacteriophage lambda DNA digested with HindIII (0.5 mg/ml, GIBCO BRL). The marker used for determining size and concentration was High DNA Mass Ladder (Invitrogen Life Technologies). All electrophoresis was carried out in 1 × TBE buffer.

2.2.9 DNA purification by the QIAquick PCR purification kit

This method was used to purify RT-PCR products prior to ligation into a vector. This was carried out according to the QIAquick PCR Purification Kit protocol with the following specifics:

- The first and second centrifugation steps were carried out for 60 sec.
- The volume of Buffer EB used was 30 µl.
- The column was allowed to stand for 1 min before the final centrifugation step.

2.2.10 DNA Purification by exonuclease I and shrimp alkaline phosphatase treatment

This method was used to purify colony PCR products prior to sequencing. The following volumes of components were mixed for each purification:

Component	Volume (μl)
PCR product	5
Exonuclease I 10 U/μl, USB	1
Shrimp alkaline phosphatase 1 U/μl, USB	2

The samples were then incubated on a Peltier Thermal Cycler DNA Engine DYAD at 37°C for 15 min, followed by 80°C for 15 min to stop the reaction.

2.2.11 Sequencing

This method was used to sequence colony PCR products. The following volumes of components were mixed on ice for each sequencing reaction:

Component	Volume (μl)
PCR product	8
dH_2O	3
Big Dye Terminator Cycle Sequencing Ready Reaction Mix Applied Biosystems	4
Diluting buffer 200 mM TRIS pH9.0 + 5 mM $MgCl_2$	4
3.2 pmol primer diluted according to oligonucleotide length in dH_2O	1

The primers used were either for the T7 or the SP6 RNA polymerase promoter and one of each was carried out for each product, to sequence the product in opposite directions. Cycle sequencing was then carried out on a Peltier Thermal Cycler DNA Engine DYAD using the following programme, with the reactions being kept on ice and placed on the block once it had reached 96°C.

Temperature (°C)	Time (min)	
96	0.50	} 24 cycles
50	0.25	
60	4.00	

Ethanol precipitation was then carried out. Each product was added to 50 μl ethanol and 2 μl 3M sodium acetate pH5.2 and incubated at room temperature for 1-8 hr. They were centrifuged for 30 min at 13,000 rpm and the supernatant was immediately removed. 200 μl of 70% ethanol was added and the products were centrifuged for 15 min at 13,000 rpm. The supernatant was removed, followed by a 30 sec centrifugation at 13,000 rpm and further extraction of any remaining liquid. After air-drying at room temperature for approximately 20 min, the samples were submitted to the MRC Human Genetics Unit sequencing service for running on an ABI Prism sequencer.

2.2.12 Ligation

Ligations were carried out according to the pGEM-T Easy Vector System protocol with the following specifics:

- The volume of PCR product used was 3 μl in each standard reaction.
- No dH_2O was used in the standard reactions.
- Reactions were either incubated at 4°C overnight or for 1 hr at room temperature.

2.2.13 Transformation

Transformations were carried out according to the Subcloning Efficiency DH5α Competent Cells protocol with the following specifics:

- Unused cells were replaced immediately at -70°C without using an intermediate dry ice or ethanol bath stage.
- 2 μl of the DNA ligation was used in each transformation.
- L-broth was used for expression, with the following recipe per litre (by MRC Human Genetics Unit technical staff):

Component	**Mass (g)**
Tryptone	10
Yeast extract	5
NaCl	10
Glucose	1

- L-agar+Amp plates were used for spreading the transformed bacteria, with the following recipe per litre (by MRC Human Genetics Unit technical staff):

Component	Mass (g)
Tryptone	10
Yeast extract	5
NaCl	10
Agar	15
Ampicillin	0.05

- Each L-agar+Amp plate was spread with 20 µl of a 50 mg/ml solution of X-Gal in N,N-DMF (Sigma), which was left to soak in for at least 1 hr before spreading bacteria.
- One 100 µl plate and one 200 µl L-agar+Amp+X-Gal plate were spread for each transformation.

2.2.14 Midiprepping

Midipreps were carried out according to the QIAGEN Plasmid Midi Protocol in the QIAGEN Plasmid Purification Handbook with the following specifics:

- The starter culture was 2 ml of L broth, with 1 µg/ml of a 50 mg/ml solution of ampicillin added, as with all L-broth+Amp media.
- The starter culture was diluted 1 in 500 by adding 50 µl to 25 ml L-broth+Amp and cultured for 16 hr.
- All centrifugation was carried out in a Sorvall Super T21 centrifuge.
- Before applying to the QIAGEN-tip, the supernatant was filtered through pre-soaked gauze.

- The purified DNA was resuspended in 400 µl TE, with the following TE composition:

Component	Concentration (mM)
TRIS HCl pH7.5	10
EDTA	1

2.2.15 Ethanol precipitation

Ethanol precipitation was carried out to increase the concentration of a DNA sample. A volume of 3 M sodium acetate pH5.2 equal to one tenth that of the sample was added, followed by a volume of ice-cold ethanol equal to 2.5 times that of the sample. This was mixed and kept at -70°C for 15 min. The sample was then centrifuged for 5 min at 13,000 rpm and the supernatant removed. One ml room temperature 70% ethanol was added and the sample was inverted. It was centrifuged for 5 min at 13,000 rpm and the supernatant was removed. The sample was then air-dried and an appropriate volume of dH_2O was added for the required concentration.

2.2.16 Restriction

All restriction enzymes were from Roche or New England Biolabs. Digests were carried out in a 40 µl volume for confirmation or quantification purposes and a 50 µl volume to produce DNA for use in synthesising *in situ* hybridisation riboprobes. The restriction enzyme was added at a concentration of 1-1.2 U/µl, with all restrictions

being incubated overnight. Guidelines provided by the manufacturer were followed to determine the correct buffer and incubation temperature for each enzyme.

2.2.17 Blunt ending

Due to the presence of an NcoI restriction site in the Mm.22913 primer amplified insert, SacII was used in a plasmid restriction for probe synthesis. This enzyme leaves a 3' overhang, therefore blunt ending was performed to allow correct probe synthesis. This was done by adding 4 µl T4 polymerase (Roche) and 5 µl each of 25mM dATP, dCTP, dGTP and dTTP to the plasmid restriction. The DNA was then incubated at 37°C for 15 min followed by incubation at 75°C for 10 min to stop the reaction.

2.2.18 Purification Using CHROMA SPIN Columns

Purification using columns was used for restriction reaction products required for synthesising *in situ* hybridisation riboprobes and, during the procedure for synthesising ^{35}S labelled riboprobes, for removal of unincorporated label. CHROMA SPIN+DEPC H_2O columns were used in the form of CHROMA SPIN-100 or CHROMA SPIN-30. Purification was carried out according to the CHROMA SPIN Columns User Manual with no alterations.

2.3 *In situ* hybridisation techniques

2.3.1 Synthesis of DIG labelled riboprobes for *in situ* hybridisation

The following volumes of components were mixed for each labelling reaction, with a total volume of 20 µl and DNA being derived from a column purified linearised plasmid:

Component	**Volume (µl)**
10 × transcription buffer Roche	2
10 × DIG labelling mix Roche	2
1 µg DNA from appropriate restriction reaction	X_1
RNase inhibitor 40 U/µl, Roche	1
DEPC dH_2O	X_2
RNA polymerase 20 U/µl, T7 or SP6 as appropriate, Roche	2

The labelling reactions were incubated at 37°C for 1 hr after which 1 µl was removed from each reaction to check by agarose gel electrophoresis, and 1 µl of the appropriate RNA polymerase was added to each reaction. A further incubation at 37°C for 1 hr was completed and 1 µl of DNase I (10 U/µl, Roche) was added to each reaction. The reactions were incubated at 37°C for 15 min before the reaction was stopped. This was done by adding the following volumes of components to each reaction:

Component	Volume (µl)
200 mM EDTA pH8	2.0
4 M lithium chloride	2.5
Ethanol	75.0

The reactions were stored at -20°C for 30 min, followed by centrifugation for 15 min at 13,000 rpm. The supernatant was removed and 200 µl 70% ethanol (in DEPC dH_2O) was added to each reaction. The reactions were centrifuged for 15 min at 13,000 rpm, the supernatant was removed and the product was resuspended in 40 µl DEPC dH_2O.

2.3.2 Synthesis of ^{35}S labelled riboprobes for *in situ* hybridisation

At least one full reaction was carried out for each antisense riboprobe and a half reaction was carried out for each sense probe, with precisely half the volume of all components being used. The following volumes of components were mixed for each full labelling reaction, with a total volume of 11 µl and with DNA derived from a column purified linearised plasmid:

Component	Volume (µl)
10 × transcription buffer	3.0
10 mM dA,C,GTP 10 µl each of 100mM dA-, dC- and dGTP plus 70 µl DEPC dH_2O	3.0
1 M DTT Melford	1.0
1 µg DNA from appropriate restriction reaction	X_1
RNase inhibitor	1.2
DEPC dH_2O	X_2

1 M DTT was made up by measuring out 7.7 g for each 50 ml required and making up to final volume using DEPC dH_2O.

10 µl of ^{35}S rUTP (>1 µCi/100 µl, NEN) and 0.8 µl of the appropriate RNA polymerase, either T7 or SP6 (20 U/µl, Roche), were added to each labelling reaction. The reactions were incubated at 37°C for 25 min, after which 0.8 µl of the appropriate RNA polymerase was added to each reaction. The reactions were incubated at 37°C for another 25 min. One µl of DNase I (10U/µl, Roche) was added to each labelling reaction and they were incubated at 37°C for 10 min. The reactions were stopped by adding 1 µl 200 mM EDTA. Unincorporated label was removed using CHROMA SPIN-100 columns.

2.3.3 Embryo treatment for wholemount *in situ* hybridisation

After dissection, embryos required for wholemount *in situ* hybridisation were stored in 4% PFA (Sigma) at 4°C overnight and dehydrated using the following washes with shaking:

Wash solution	**Time (min)**
PBS	5
PBS	5
25% methanol/PBST	5
25% methanol/PBST	X
50% methanol/PBST	5
50% methanol/PBST	X
75% methanol/PBST	5
75% methanol/PBST	X
100% methanol/PBST	5
100% methanol/PBST	X

Each litre of PBST was made up by including 5 ml 10% Tween 20 (Sigma, made up in DEPC dH_2O) in a litre of PBS.

The X value varies according to the embryo stage as follows:

Embryo stage	X value
E9.5	25
E10.5	40
E11.5	50
E12.5	60

2.3.4 Tissue treatment for paraffin wax embedding and sectioning

After dissection, embryos required for wholemount *in situ* hybridisation were stored in 4% PFA at 4°C overnight. The following day, all tissue was dehydrated using the following washes:

Wash solution	Time (min)
PBS	10
PBS	10
10% ethanol/PBS	30
30% ethanol/PBS	30
50% ethanol/PBS	30

For all embryo stages other than E9.5, the tissue was removed from 50% ethanol/PBS and stored in 70% ethanol. Further processing was carried out by a Tissue-tek VIP, with all washes lasting 1 hr with the exception of the variable 70% ethanol/PBS wash, before embedding in paraffin wax:

Wash solution	Temperature (°C)
70% ethanol/PBS	40
85% ethanol/PBS	40
95% ethanol/PBS	40
100% ethanol/PBS	40
100% ethanol/PBS	40
xylene	40
xylene	40
paraffin wax	60
paraffin wax	60
paraffin wax	60
paraffin wax	60

For E9.5 embryos, the following washes were carried out by hand, with all washes lasting 30 min, before embedding in paraffin wax:

Wash solution	Temperature (°C)
70% ethanol/PBS	40
85% ethanol/PBS	40
95% ethanol/PBS	40
100% ethanol/PBS	40
100% ethanol/PBS	40
100% ethanol/PBS	40
xylene	40
xylene	60
paraffin wax	60
paraffin wax	60
paraffin wax	60
paraffin wax	60

Sectioning was carried out using a Reicheit-Jung Biocut 2030 microtome set at 7 μm, with sections floated out in a 42°C waterbath before mounting on polysine or Superfrost Plus slides (BDH Laboratory Supplies). Slides were incubated at 37°C overnight before storing at room temperature.

2.3.5 Tissue treatment for OCT embedding and cryostat sectioning

After dissection, all tissues were embedded in OCT (Sakura) immediately by placing in a mould containing OCT, which was then placed in a weighing boat of isopentane on dry ice. The samples were stored at -70°C until they were sectioned. A Jung Frigocut Cryostat was used for sectioning, with the box temperature at -18°C and the object temperature at -13°C. Ten μm sections were mounted on Superfrost Plus slides and air dried for a minimum of 30 min before carrying out fixation, a maximum of 4 hr after sectioning. The following washes were used in the fixation process:

Wash solution	**Time (min)**
4% PFA	5.0
PBS	2.5
GibcoBRL	
PBS	2.5
70% ethanol/DEPC dH_2O	5.0
95% ethanol/DEPC dH_2O	5.0

The slides were stored in 95% ethanol/ DEPC dH_2O at 4°C until required.

Embryos for sectioning that had undergone wholemount *in situ* hybridisation were washed in PBS, stored in 5% sucrose overnight and washed in PBS immediately before embedding in OCT.

2.3.6 Wholemount *in situ* hybridisation

Embryos for wholemount *in situ* hybridisation were pre-treated by taking them through the following washes:

Wash solution	**Time (min)**
75% methanol/PBST	10
50% methanol/PBST	10
25% methanol/PBST	10
PBST	10
PBST	10
PBST	10
6% H_2O_2 30% H_2O_2 (Sigma) diluted in PBST	60
Proteinase K 10 mg/ml stock diluted in PBST to 1 μg/ml	X
Glycine/PBST Glycine from Sigma, made up to 2 mg/ml in PBST	5
PBST	10
PBST	10
0.2% GTA/4% PFA/PBST GTA from Sigma	20
PBST	10
PBST	10

The X value for proteinase K wash time varies according to the embryo stage as follows:

Embryo stage	**X value**
E9.5	5
E10.5	9
E11.5	12
E12.5	15

Hybridisation mix was made up with the following amounts of components per 5 ml:

Component	Amount
Formamide (ultrapure) Sigma, deionised by stirring with 5 g monobed resin per 2.5l for 4 hr and filtered	2.5 ml
20 × SSC pH8.0	1.25 ml
5 mg/ml heparin Sigma	50 µl
10% Tween 20	50 µl
Blocking agent Roche	50 mg
10% CHAPS Melford Laboratories, 5 g in 50 ml DEPC dH_2O	250 µl
0.2 M EDTA pH8.0 BDH, 7.44 g in each 100 ml made up with DEPC dH_2O	250 µl
10 µg/ml tRNA Sigma	50 µl
DEPC dH_2O	600 µl

One litre of 20 × SSC contained 175.3 g NaCl and 88.24 g Tri-sodium citrate made up in dH_2O.

The pre-treated embryos were sunk in 0.5 ml hybridisation mix contained in a 5 ml cryotube. The hybridisation mix was removed and replaced with fresh mix and the embryos were incubated at 65°C for 2 hr with shaking. DIG labelled riboprobes were diluted 1:200 in hybridisation mix and denatured at 80°C for 7.5 min before placing briefly on ice. Hybridisation mix was completely removed from the cryotubes containing the embryos and replaced with hybridisation mix containing probe. The embryos were incubated overnight at 65°C with shaking.

The following post-hybridisation washes were carried out on the embryos at 65°C with shaking:

Wash solution	Time (min)
Solution A	2
Solution A	5
70% solution A in DEPC dH_2O	5
30% solution A	5
2 × SSC/0.1% CHAPS	30
2 × SSC/0.1% CHAPS	30
0.2 × SSC/0.1% CHAPS	30
0.2 × SSC/0.1% CHAPS	30

Solution A was made up with the following volumes of components per 100 ml:

Component	Volume (ml)
20 × SSC	5.0
Deionised formamide Sigma	50.0
dH_2O	43.9
Tween 20	0.1

The following washes were carried out at room temperature with shaking:

Wash solution	Time (min)
TBST	10
TBST	10
TBST	10
10% HISS/TBST block HISS from Diagnostics Scotland	60

10 × TBST was made up in DEPC dH_2O with the following components per 500 ml:

Component	Amount
NaCl Sigma	40 g
KCl BDH	1 g
1 M TRIS pH7.5	125 ml
10% Tween 20	50 ml

1 M TRIS pH7.5 contained 12.1 g TRIS base (Sigma) in each 100 ml.

The 10% HISS/TBST block was removed from the embryos and replaced with anti-DIG AP/HISS/TBST, which was made by diluting anti dig antibody (Roche) 1:2000 in 10% HISS/TBST. The embryos were incubated at 4°C overnight.

The following washes were carried out in the substrate reaction:

Wash solution	Time (min)
NTMT	10
NTMT	10
NTMT	10
NTMT	10
Substrate 35 µl BCIP (Roche) and 45 µl NBT (Roche) added to 10 ml NBT	X

NTMT (alkaline phosphate buffer) was made up in DEPC dH_2O with the following components per 100 ml:

Component	Amount
5 M NaCl	2 ml
1 M $MgCl_2$	5 ml
1 M TRIS pH9.5	10 ml
10% Tween 20	1 ml
Levamisole Vector	20 drops

X varied according to the time taken for an expression pattern to develop. The following washes were carried out to stop the substrate reaction:

Wash solution	Time (min)
STOP buffer	5
STOP buffer	5
STOP buffer	5

STOP buffer was made up with 75 ml PBS and 25 ml 0.2 M EDTA for each 100 ml. Embryos were stored in 4% PFA at 4°C.

2.3.7 Cryosection *in situ* hybridisations using DIG labelled riboprobes

Cryosection slides were removed from 4°C in 95% ethanol/DEPC dH_2O and the following washes were carried out:

Wash solution	Time (min)
70% ethanol/ DEPC dH_2O	2
50% ethanol/ DEPC dH_2O	2
30% ethanol/ DEPC dH_2O	2
PBS	10
PBS	10
Proteinase K 10 mg/ml stock diluted in proteinase K buffer to 500 ng/ml	10
PBS	2
PBS	2
0.1 M TEA and acetic anhydride 625 µl acetic anhydride per 200 ml 0.1 M TEA	5
0.1 M TEA and acetic anhydride	5
1 × salts	2

Proteinase K buffer was composed of 25 ml 1 M TRIS pH7.5 and 12.5 ml 0.2 M EDTA made up in DEPC dH_2O for each 500 ml. 0.1 M TEA pH8 was made by using 18.6 g TEA in DEPC dH_2O for each litre. The 1 × salts was diluted from a 10 × salts pH7.5 stock solution containing the following amounts of components:

Component	**Amount**
NaCl	114 g
TRIS HCl	14 g
TRISbase	1.34 g
NaH_2PO_4	7.8 g
Na_2HPO_4	7.1 g
0.5 M EDTA	100 ml

Hybridisation mix was made up with the following amounts of components per ml:

Component	**Amount (µl)**
Formamide (ultrapure)	500
50% dextran sulphate Sigma, 500 µg/ml in DEPC H_2O	200
50 × Denhardts Sigma, 150 mg in 5 ml DEPC H_2O	20
1 M TRIS pH8	20
5 M NaCl	60
0.2 M EDTA	25
0.1 M $NaPO_4$ BDH Laboratory Supplies	100
10 µg/ml tRNA	50
DEPC dH_2O	25

DIG labelled riboprobes were diluted in hybridisation mix 1 in 200 and denatured at 80°C for 5 min. One hundred µl of the relevant hybridisation mix was overlaid on each slide using Hybrislips (Surgipath). The slides were incubated at 55°C overnight in a slide box humidified with 50% formamide/salts. After overnight hybridisation, the following washes were carried out on the slides:

Wash solution	Time (min)	Temperature
Solution A	15	65°C
Solution A	30	65°C
Solution A	30	65°C
TBST	30	room temperature
TBST	30	room temperature

Each 250 ml of solution A was composed of 12.5ml 20 × SSC, 125 ml formamide and made up with dH_2O before adding 250 µl 10% Tween 20.

The slides were incubated in 10% HISS/TBST at room temperature for more than an hour followed by overnight incubation at 4°C with anti-DIG-AP in HISS/TBST. This overnight incubation was carried out by either overlaying the slides with the antibody using glass coverslips, placing the slides in a container humidified with TBST, or by submerging the slides in the antibody. The following washes were carried out after overnight incubation:

Wash solution	Time (min)
TBST	20
TBST	20
TBST	20
TBST	20
NTMT	10
NTMT	10
Substrate	X

The X value varied according to the time taken for an expression pattern to develop. The slides were washed twice in water and incubated in 4% PFA for a minimum of 20 min and overnight at the most. The slides were then washed twice in water before staining with filtered Nuclear Fast Red (Vector), washing in water and carrying out the following washes:

Wash solution	Time (min)
30% ethanol/ dH_2O	0.5
50% ethanol/ dH_2O	0.5
70% ethanol/ dH_2O	0.5
90% ethanol/ dH_2O	0.5
100% ethanol	5.0
100% ethanol	5.0
Histoclear	5.0
National Diagnostics	
Histoclear	5.0

The slides were then mounted using Histomount (National Diagnostics).

2.3.8 Wax section *in situ* hybridisations using ^{35}S labelled riboprobes

The following washes were carried out on wax section slides:

Wash solution	**Time**
Xylene	5 min
Xylene	5 min
100% ethanol	2 min
100% ethanol	2 min
100% ethanol	2 min
90% ethanol/ DEPC dH_2O	2 min
70% ethanol/ DEPC dH_2O	2 min
50% ethanol/ DEPC dH_2O	2 min
30% ethanol/ DEPC dH_2O	2 min
PBS	2 min
4% PFA	10 min
PBS	2 min
PBS	2 min
Proteinase K	7.5 min
10 mg/ml stock diluted in proteinase K buffer to 500 ng/ml	
PBS	1 min
4% PFA	2 min
DEPC dH_2O	10 sec
0.1 M TEA	30 sec
0.1 M TEA and acetic anhydride	5 min
625 µl acetic anhydride per 200 ml 0.1 M TEA	
0.1 M TEA and acetic anhydride	5 min
PBS	2 min
0.85% NaCl	2 min
Sigma	
30% ethanol/ DEPC dH_2O	1 min
50% ethanol/ DEPC dH_2O	1 min
70% ethanol/ DEPC dH_2O	1 min
90% ethanol/ DEPC dH_2O	1 min
100% ethanol	5 min
100% ethanol	5 min
100% ethanol	5 min

Slides were air dried before hybridisation. Hybridisation mix was made up with the following amounts of components per ml:

Component	Amount (µl)
Formamide (ultrapure)	500
50% dextran sulphate	200
50 × Denhardts	20
1 M TRIS pH8	20
5 M NaCl	60
0.2 M EDTA	25
0.1 M $NaPO_4$	100
10 µg/ml tRNA	50
1 M DTT	25
7.7 g in 50 ml DEPC H_2O	

A tenth of the final hybridisation mix consisted of the appropriate volume of probe to give 7.2 µCi per slide, with DDT in DEPC H_2O making up the difference. Probes were denatured at 80°C for 2 min and 80 µl of the relevant hybridisation mix was overlaid on each slide using Hybrislips. The slides were incubated at 55°C overnight in a slide box humidified with 50% formamide/5 × SSC. After overnight hybridisation, the following washes were carried out on the slides:

Wash solution	Time	Temperature
5 × SSC	20 sec	55°C
5 × SSC/DTT 2.5 ml 1 M DTT and 62.5 ml 20 × SSC in 250 ml H_2O	10 min	55°C
High stringency wash	30 min	65°C
NTE	10 min	37°C
NTE	10 min	37°C
NTE	10 min	37°C
NTE/RNase A 1 ml of 10 mg/ml RNase A per 600 ml NTE	30 min	37°C
NTE	5 min	37°C
NTE	10 min	37°C
2 × SSC	15 min	room temperature
2 × SSC	15 min	room temperature
2 × SSC	15 min	room temperature
2 × SSC	15 min	room temperature
0.1 × SSC	15 min	room temperature
0.1 × SSC	15 min	room temperature
0.1 × SSC	15 min	room temperature
0.1 × SSC	15 min	room temperature
30% ethanol and 3 M NH_4OAc 200 ml ethanol mix and 20 ml NH_4OAc	1 min	room temperature
50% ethanol and 3 M NH_4OAc	1 min	room temperature
70% ethanol and 3 M NH_4OAc	1 min	room temperature
90% ethanol and 3 M NH_4OAc	1 min	room temperature
100% ethanol	5 min	room temperature
100% ethanol	5 min	room temperature

Slides were air dried and left on film (Kodak) for a week before developing the film and dipping the slides in K5 emulsion (Ilford) under safe light conditions. Emulsion was incubated at 42°C for 14 min before diluting in an equal volume of 1% glycerol and incubating at 42°C for a further 2 min. The slides were then dipped and left to dry in light tight containers overnight before boxing them and storing at 4°C for 4-10 weeks until ready to be developed. The slides were left at room temperature for at least an hour prior to developing by carrying out the following washes under safelight conditions:

Wash solution	Time
D19	4 min
Kodak	
H_2O	10 sec
Fix	5 min
diluted 1:3 in H_2O, Amfix	
H_2O	5 min
H_2O	5 min
H_2O	10 min

The slides were then counterstained with 1% methyl green and mounted using DePex (BDH Laboratory Supplies).

2.4 Microscopy

Embryo imaging was performed with a Leica Stereo MZFLIII stereo fluorescence microscope (Leica Microsystems) fitted with a fibre optic cold light source and illuminated base for incident/transmission illumination. Image acquisition was with a Photometrics CoolSnap colour CCD camera (Roper Scientific) controlled by scripts written for IPLab Spectrum (Scanalytics Inc).

Section imaging was performed with Zeiss Axioplan II fluorescence microscope (Carl Zeiss Ltd) equipped with colour additive filters (Andover Corp) for sequential colour imaging and a Photometrics CoolSnap HQ monochrome CCD camera (Roper Scientific). Image capture was by scripts written for IPLab Spectrum (Scanalytics Inc) which controlled camera capture and filter selection via motorised filter wheels (Ludl Electronic Products Ltd).

CHAPTER 3

Digital Differential Display

3.1 Introduction

The availability of publicly accessible data sets and bioinformatic tools for their manipulation have made gene discovery a faster and easier process than was previously possible. Differential expression analysis is a useful method to elucidate genes that have an important input in the development or function of an organ, tissue or cell type of interest. In this thesis, an *in silico* differential display approach was taken to facilitate the discovery of novel genes that have enriched expression in the eye compared with other organs and tissues. To this end, the UniGene database was used with the accompanying programme Digital Differential Display (DDD).

UniGene is a database that contains sequences from a variety of species including human and mouse (Boguski and Schuler, 1995). The majority of these sequences are ESTs deposited in dbEST and deriving from specified cDNA libraries. In an effort to reduce redundancy in the raw data set and allow its use to be effective, the ESTs from a species are clustered, with the clusters representing putative genes and, in some cases, known genes. The first step in the build procedure screens the ESTs and ensures removal of repetitive, low complexity and contaminant sequences, subsequently removing those less than 100bp in length. The set of mRNA sequences from GenBank is clustered before creating initial links with ESTs. The links joining mRNA clusters to each other are severed and the EST clusters are built up as similarities between them are found. Clone-based edges are then used to cluster non-overlapping 5' and 3' ESTs together, with stringency provided by the necessity of at least two 5' ESTs from one cluster with at least two 3' clusters from the other before they are merged. At this stage unanchored clusters, that is those without a sequence containing a polyadenylation signal or tail, are eliminated. Unclustered ESTs are then rechecked for similarity at a lower stringency than before and those meeting the requirements are joined to the relevant cluster. A similar process is applied to clusters with one EST, with comparisons being made with other clusters at lower stringency than previously and mergers taking place if appropriate. Every new build is compared with the last to provide continuity in annotation. Each cluster is given an

identifier, although these are not permanent, as clusters and the sequences within them can be removed or added to as more sequences are entered into UniGene. (Pontius et al., 2003)

DDD (http://www.ncbi.nlm.nih.gov/UniGene/info_ddd.shtml) is able to measure the number of sequences from a cluster that appear in a defined set of libraries. The libraries available for use in DDD analysis are those with over 1,000 sequences in UniGene, which provides a high enough number for statistical significance to be shown for the frequency of gene representation. Within DDD, pools of two or more libraries can be constructed by the user from these available libraries. For each cluster within UniGene, the number of ESTs that have derived from each pool of libraries is represented as a proportion of the total number of ESTs deriving from that pool. The output of DDD consists of clusters with a statistically significant difference between the EST representations of the two pools. It can therefore be utilised as a resource for the identification of genes that may contribute to the identity of a specific cell type, tissue or organ. (Pontius et al., 2003)

DDD was used in preference to utilising the search for tissue specific transcripts available at TIGR (http://www.tigr.org/tdb/tgi/) to mine their Gene Index databases. The Gene Index databases have similar content to the UniGene databases in that there are clusters of ESTs, known as tentative consensuses (TCs) at TIGR. As suggested by the name, these clusters are unlike those at UniGene, as the ESTs within them must be overlapping by over 40 bases and have at least 95% percentage identity in these regions. This implies a greater degree of redundancy in the Gene Index databases than in UniGene, therefore UniGene will give a greater accuracy with respect to clusters directly representing a gene (Schuler, 1997; Bouck et al., 1999).

In this thesis, the DDD procedure was used multiple times. The initial DDD was carried out on mouse build 82. The second DDD produced the results that formed the basis for further experimental investigation and utilised mouse build 100. The third and final DDD to be carried out on the mouse UniGene set was on build 120 and

illustrates the current situation of UniGene. Human build 160 was used for a DDD to allow comparison with the results derived from mouse and confirmation of the preferential expression shown of mouse clusters by analysing their human orthologues. Finally, a rat DDD on build 119 was carried out. This was also an attempt to confirm preferential eye expression in orthologues of mouse clusters, but instead demonstrated the effects of bias in tissue composition of pool libraries. As well as analysing the output of DDD at the time of initiation, analysis has also been carried out on the results in the light of current available information on the relevant clusters in order to gain a more complete and accurate analysis. This serves to highlight the shifting nature of the UniGene databases.

3.2 Mouse Digital Differential Display: build 82

3.2.1 Composition of mouse build 82

The first DDD was performed in October 2000 on mouse build 82. The sequence composition of this build is shown (Table 3.1). The majority of sequences are 3' ESTs (68%), with 5' ESTs making up most of the remaining total (29%). Unclassified ESTs and mRNAs plus coding sequences are a minor contribution to the build in numerical terms (2% and 1% respectively).

Sequence type	Number of sequences				
	Mm 82	Mm 100	Mm 120	Hs 160	Rn 119
mRNAs + gene CDSs	17,508	38,576	47,518	106,900	13,685
HTC	-	-	53,905	7,500	5
EST, 3' reads	980,395	1,254,576	1,575,035	1,451,825	261,813
EST, 5' reads	422,268	832,797	1,542,711	1,886,373	145,550
EST, other/unknown	26,806	104,901	223,213	682,624	5,059

Table 3.1 The sequence composition of builds used in DDD
The build organism and number are indicated. Mm, *Mus musculus*; Hs, *Homo sapiens*; Rn, *Rattus norvegicus*; CDS, coding sequence; HTC, high-throughput cDNA.

There were 72,651 clusters in mouse build 82, with 6,979 (10%) containing at least one known gene sequence, 71,954 (99%) including at least one EST and 6,282 (9%) incorporating both types of sequence. The size of clusters ranges from one to more than 4,097 sequences and the distribution of these is shown (Figure 3.1). There are 33,941 (47%) clusters in the build that can be described as singletons, in that they consist of a single sequence. The remaining clusters take on a general trend of decreasing frequency with increasing cluster size, with the exception of the 3-4 sequences bracket, which is the second most prevalent.

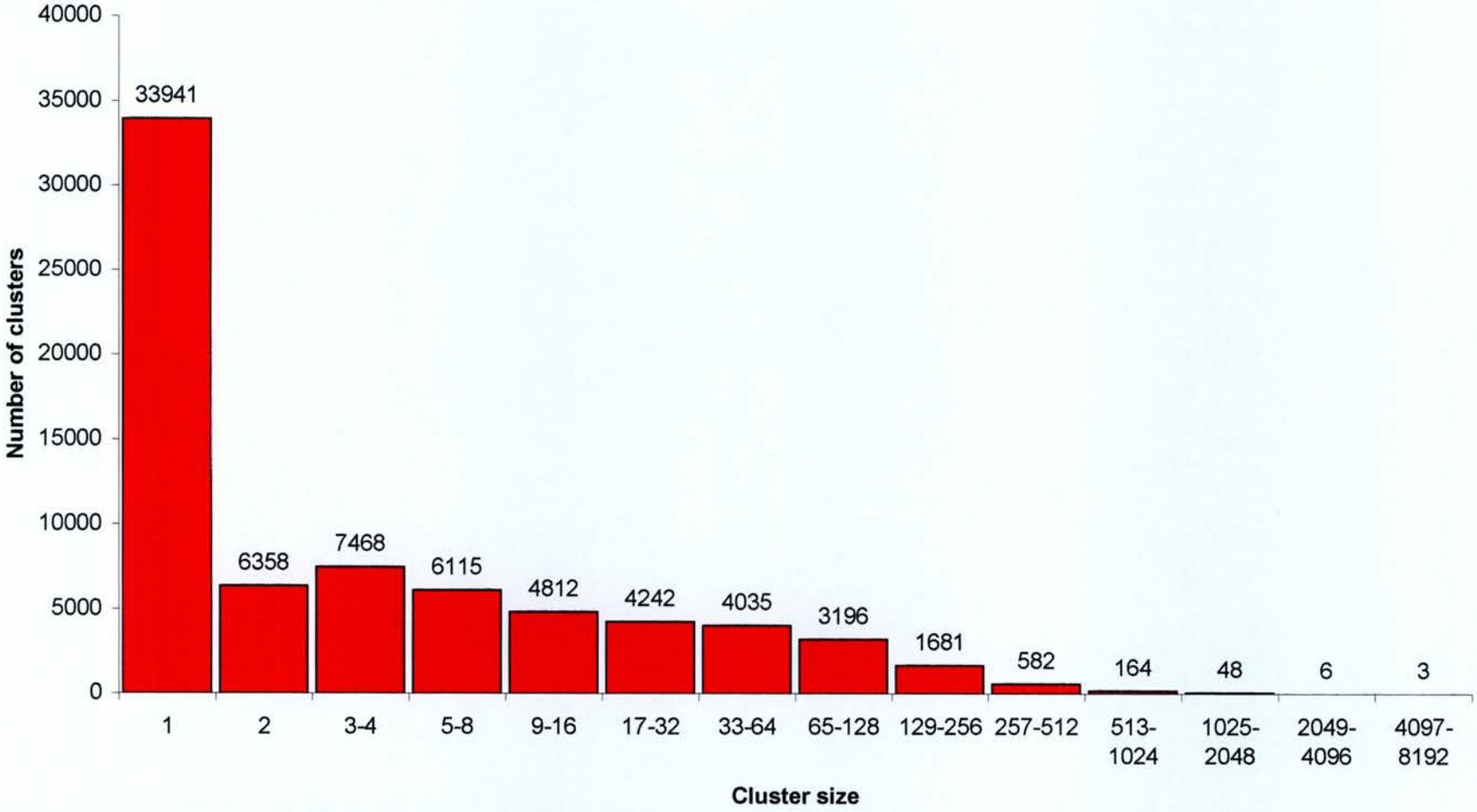


Figure 3.1 Cluster size distribution in mouse UniGene build 82
Cluster size classes are in terms of the number of sequences contained in them and the number of clusters that fall into each class are shown with the relevant number above each bar.

3.2.2 Library pool construction for DDD on mouse build 82

DDD required the construction of two pools of libraries. The first was designated the eye pool, containing the six available eye derived libraries (E9.5 optic vesicle; E11.5 optic cup; E12 eyeball; neonate eyeball; adult eyeball; adult retina) and consisting of 26,029 EST sequences in total. The second target pool comprised all other available libraries produced from normal tissues (those that are not cancerous or diseased) and organs, thereby excluding cell lines, but also omitting those libraries that would possess eye ESTs such as head and whole embryo libraries. This pool encompassed 222 libraries with 1,129,506 ESTs.

3.2.3 Original results analysis for DDD on mouse build 82

As DDD output consists of clusters with statistically significant differences, the results included clusters with higher representation in the target pool than the eye pool, that is, clusters that have lower expression in the eye than in other tissues and organs. These were removed to give a set of clusters in which the EST composition was significantly higher in the eye pool than in the target pool. The resulting clusters were divided into known and unknown subsets, with the clusters containing a known gene sequence, and thus being assigned that gene name, and the unknown subsets of clusters representing putative novel genes. In order to gain an indication of the proportion of false positives in the unknown set of clusters, the known clusters were divided into those that are eye relevant and those that have unknown eye relevance. Eye relevance was established by the presence of eye phenotypes incurred through induced or existing mutations in the gene, previous evidence of elevated representation of the gene in the eye by *in situ* expression analysis, a specific defined function of the gene within the eye or a combination of these factors. A diagram

representing the results of this analysis is shown (Figure 3.2). There are 107 (65%) clusters out of the total 165 that are unknown and may represent novel genes with eye specific or eye enriched expression. Of the remaining 58 known clusters, 33 (20% of the total; 57% of known) clusters are eye relevant.

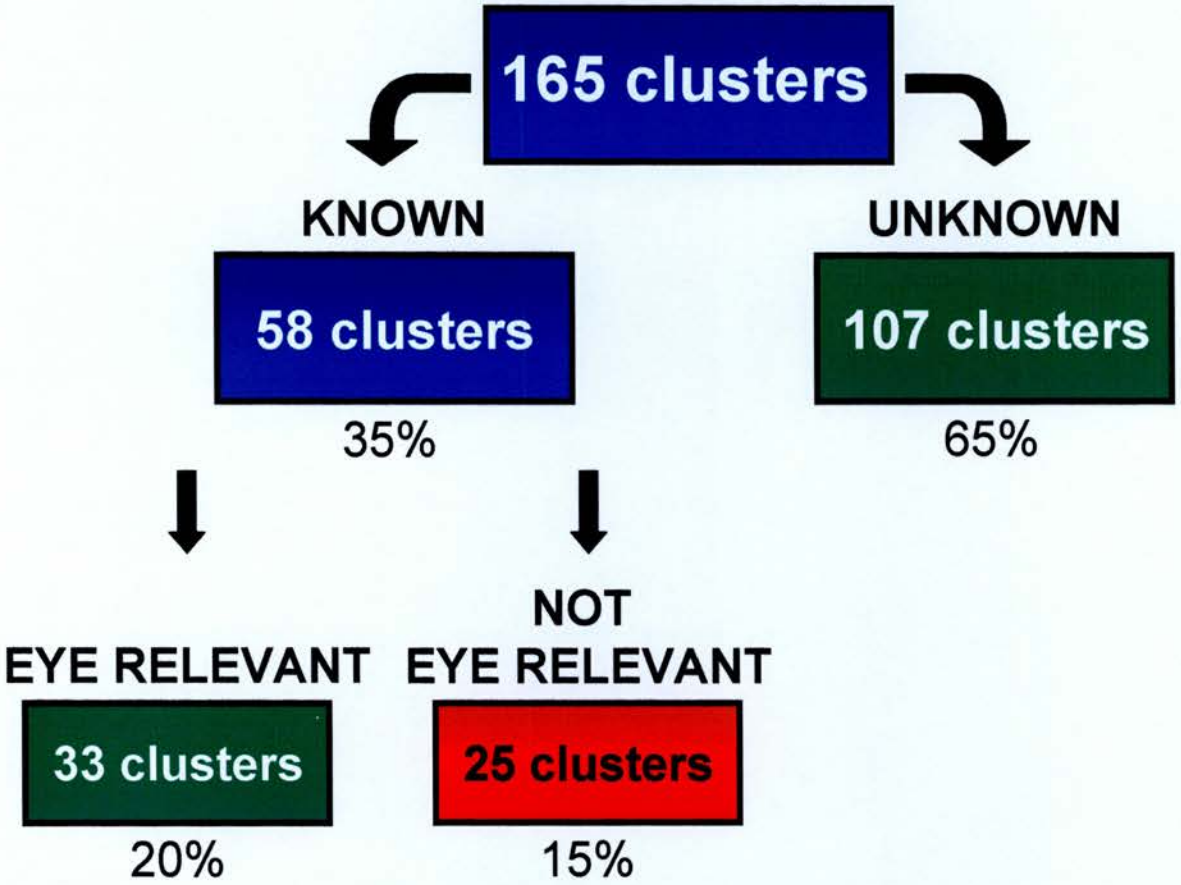


Figure 3.2 Cluster composition of mouse UniGene build 82 DDD results
The total number of clusters with significantly more expression in the eye pool was 165. The number of clusters is shown in the boxes and the percentage of the total for each category is given underneath. Known clusters are those assigned with a gene name. Eye relevant clusters are those that have been previously reported to have a specific eye expression or function.

3.2.4 Retrospective results analysis for DDD on mouse build 82

The current mouse build (120) was used to look at the results retrospectively, as some unknown clusters may have subsequently been characterised. Alternatively, unknown clusters may have been retired, with the majority of their sequences joining those in a different cluster. Of the 107 unknown clusters from the first analysis, only 36 clusters remain unretired and unknown. Of the rest, 18 are now known, with six of these being eye relevant. The remaining 53 clusters have been retired, with 22 disappearing completely and the other 31 adopted by other clusters. Six of these new clusters are eye relevant, 10 are not known to be eye relevant and 15 are unknown. Incorporating these changes, the new analysis of the results is shown (Figure 3.3). The proportion of known genes is now higher at 64%. Of these, 45 (31% of the total; 49% of known) clusters are eye relevant. The number of eye relevant clusters and their percentage of the total have increased from the original analysis, however the relative number of eye relevant clusters to those that are not known to be eye relevant is lower.

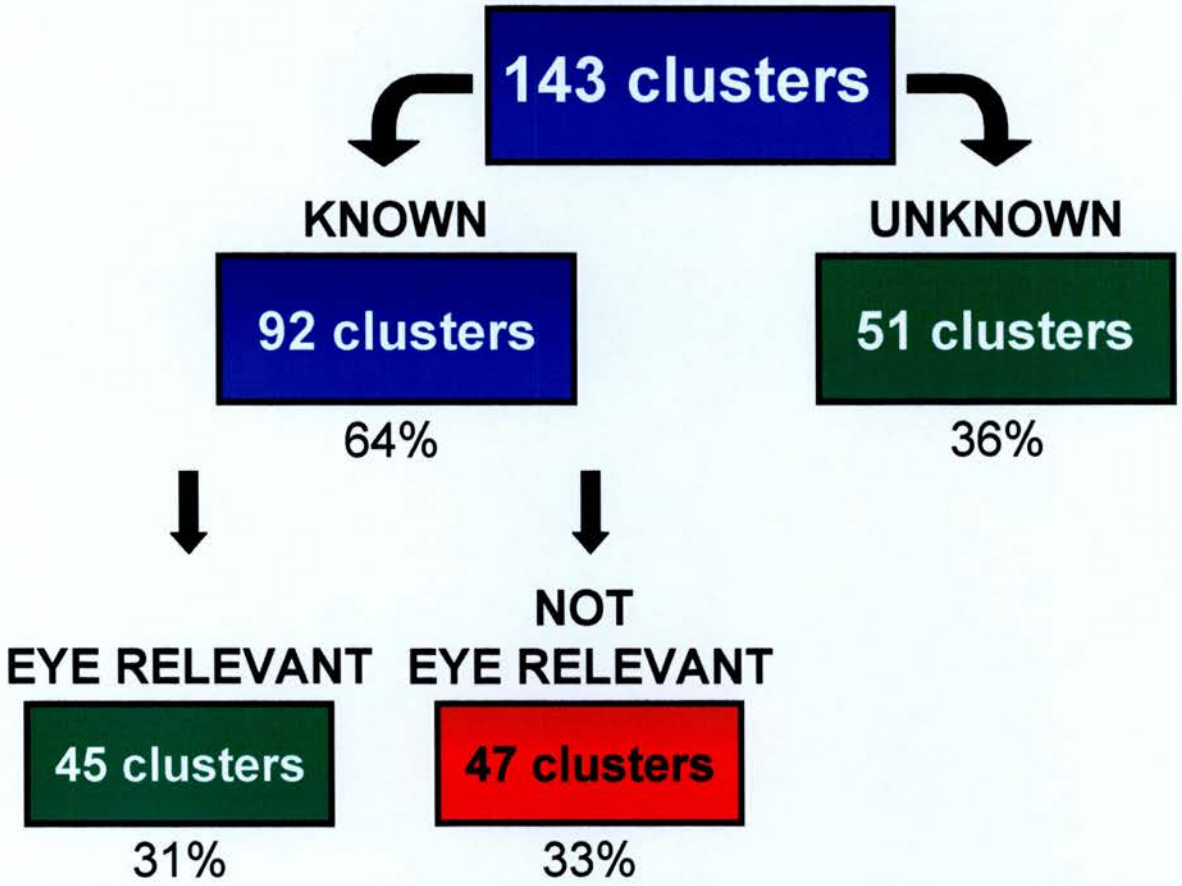


Figure 3.3 Cluster composition of mouse UniGene build 82 DDD results after applying new knowledge from build 120
The total number of clusters with significantly more expression in the eye pool that were retained intact or had sequences adopted by other clusters was 143. The number of clusters is shown in the boxes and the percentage of the total for each category is given underneath. Known clusters are those assigned with a gene name. Eye relevant clusters are those that have been previously reported to have a specific eye expression or function.

3.3 Mouse Digital Differential Display: build 100

3.3.1 Composition of mouse build 100

The second DDD was carried out in January 2002 on mouse build 100. The sequence composition of this build is shown (Table 3.1). As in mouse build 82, the majority of sequences are 3' ESTs (56%), with 5' ESTs remaining the second most represented type of sequence (37%). The increase in percentage from mouse build 82 in number of sequences is greatest for unclassified ESTs (from 2% to 5%), followed by mRNA and coding sequences (from 1% and 2%), the two categories with the smallest contributions to the total number of sequences.

The sequences in mouse build 100 were assembled into 85,049 clusters, with 16,994 (20%) including at least one gene sequence, 83,640 (98%) containing at least one EST sequence and 15,585 (18%) clusters comprising both. The size of clusters ranged from one to more than 8,193 sequences and the distribution of these is shown (Figure 3.4). In this build, 35,935 (40%) clusters are singletons, which is a decrease in percentage even though it is an increase in the number of clusters. As with build 82, the remaining clusters take on a general trend of decreasing frequency with increasing cluster size, although there is more deviation from this pattern than there was in build 82.

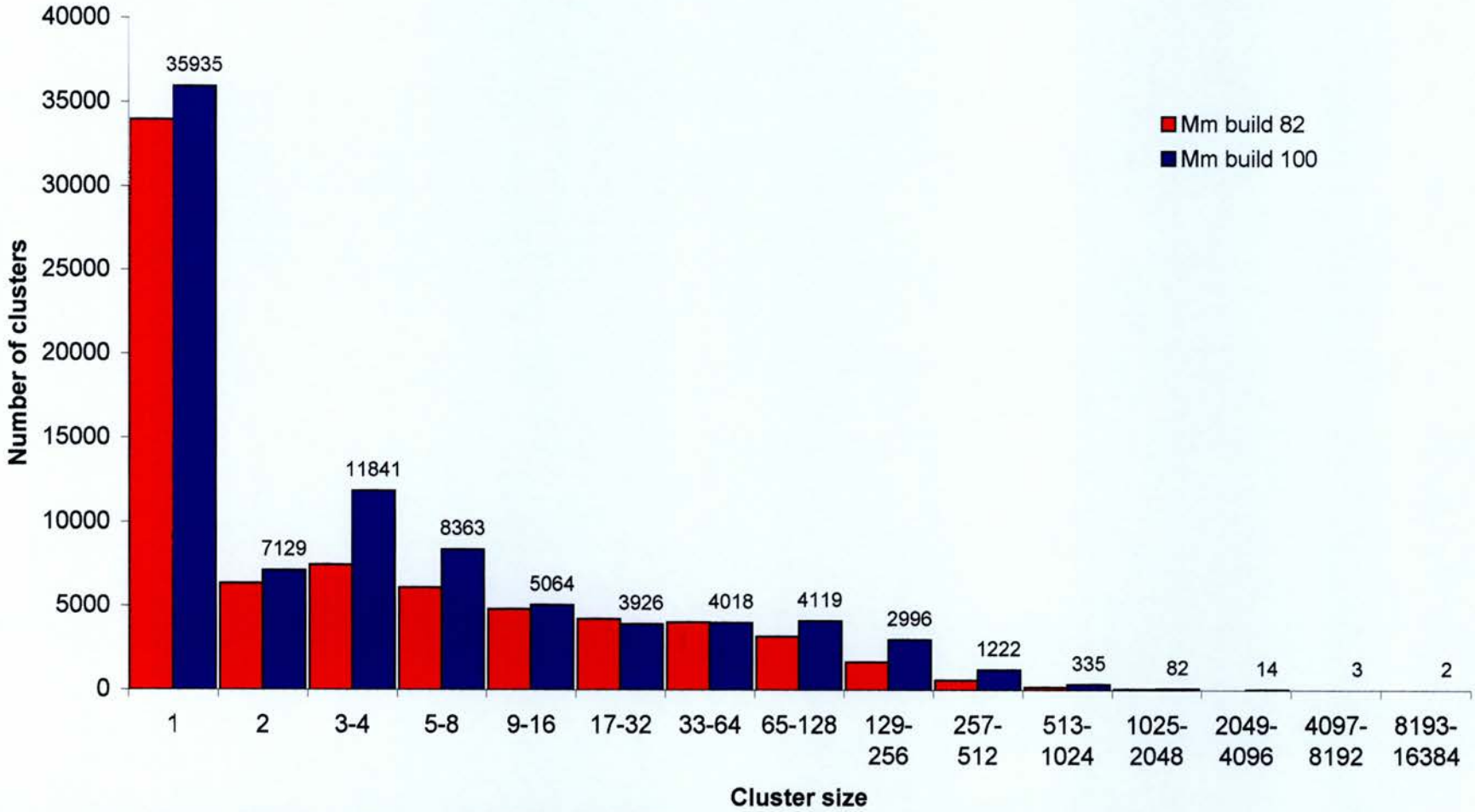


Figure 3.4 Cluster size distribution in mouse UniGene build 100
Cluster size classes are in terms of the number of sequences contained in them and the number of clusters that fall into each class are shown with the relevant number above each bar (blue). The distribution of mouse UniGene build 82 is also shown (red).

3.3.2 Library pool construction for DDD on mouse build 100

The eye pool for the DDD on mouse build 100 consisted of seven libraries (E12 eyeball; E13.5 retina; E14.5 retina; neonate eyeball; adult eyeball; two adult retina) containing 47,590 ESTs. The target pool encompassed 174 libraries with 1,433,821 ESTs.

3.3.3 Original results analysis for DDD on mouse build 100

The results were analysed in the same way as for the DDD on mouse build 82 and this is shown (Figure 3.5). Of the 292 clusters with significantly higher representation in the eye pool than in the target pool, 154 (53%) clusters are known, which is higher than the 58 (35%) clusters from the mouse build 82 DDD results. The ratio of eye relevant genes to genes with unknown eye relevance is approximately equal, with the percentage of the total (25% and 28% respectively) and the number of clusters (74 and 80 respectively) higher than in the previous DDD build due to the increase in number and proportion of known genes.

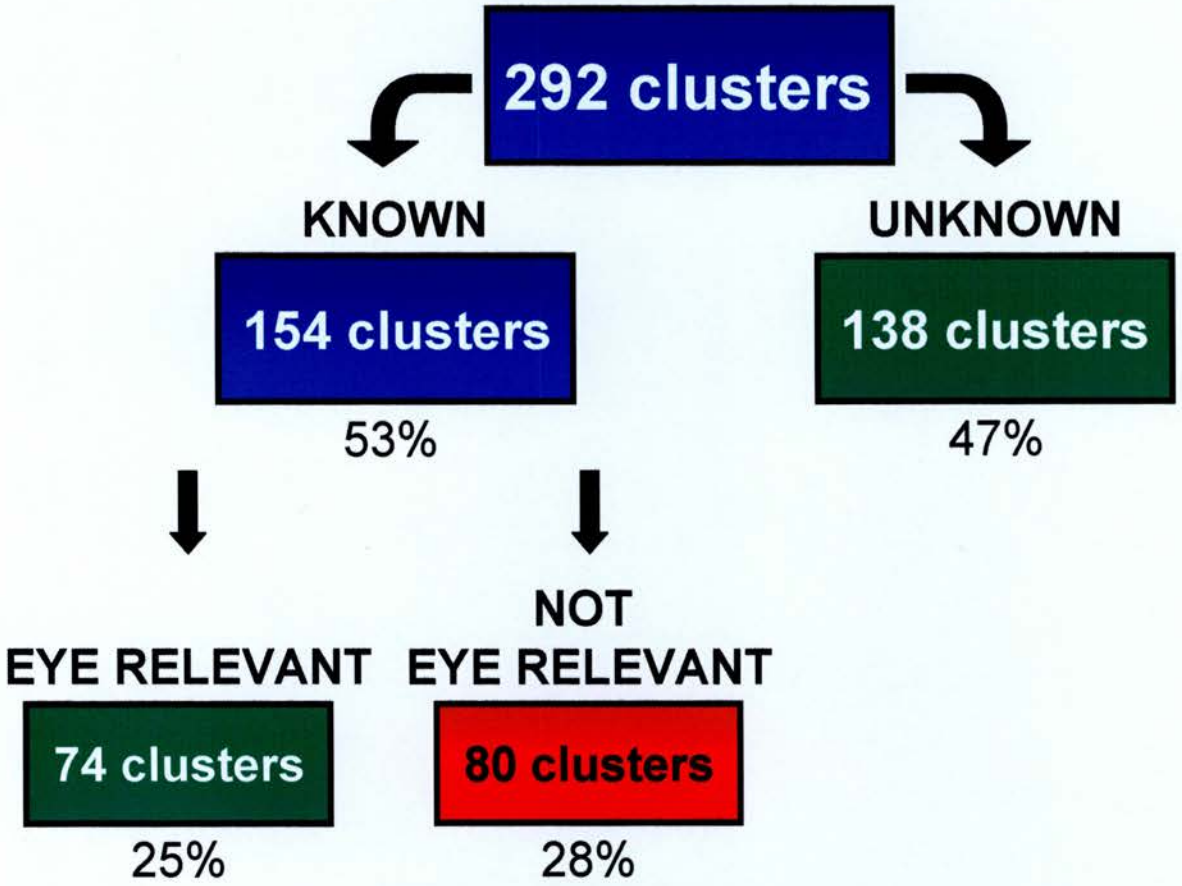


Figure 3.5 Cluster composition of mouse UniGene build 100 DDD results
The total number of clusters with significantly more expression in the eye pool was 292. The number of clusters is shown in the boxes and the percentage of the total for each category is given underneath. Known clusters are those assigned with a gene name. Eye relevant clusters are those that have been previously reported to have a specific eye expression or function.

3.3.4 Retrospective results analysis for DDD on mouse build 100

New knowledge was applied to the results from mouse build 120, as was done with those from mouse build 82. Of the 138 unknown clusters, 61 are now known, with 10 of these having eye relevance. Thirty-seven of the 138 unknown clusters were retired, with just one being entirely removed and 36 being adopted by other clusters. Of these 36, half are unknown and six of the 18 known clusters are eye relevant. The new analysis is shown (Figure 3.6). As with the earlier DDD, the percentage of known genes has increased since the previous used build, this time from 53% to 67%. The ratio of genes that are not known to be relevant to those that are is higher in the new than in the original analysis, with just 30% of the total being eye relevant and 37% not classified as such.

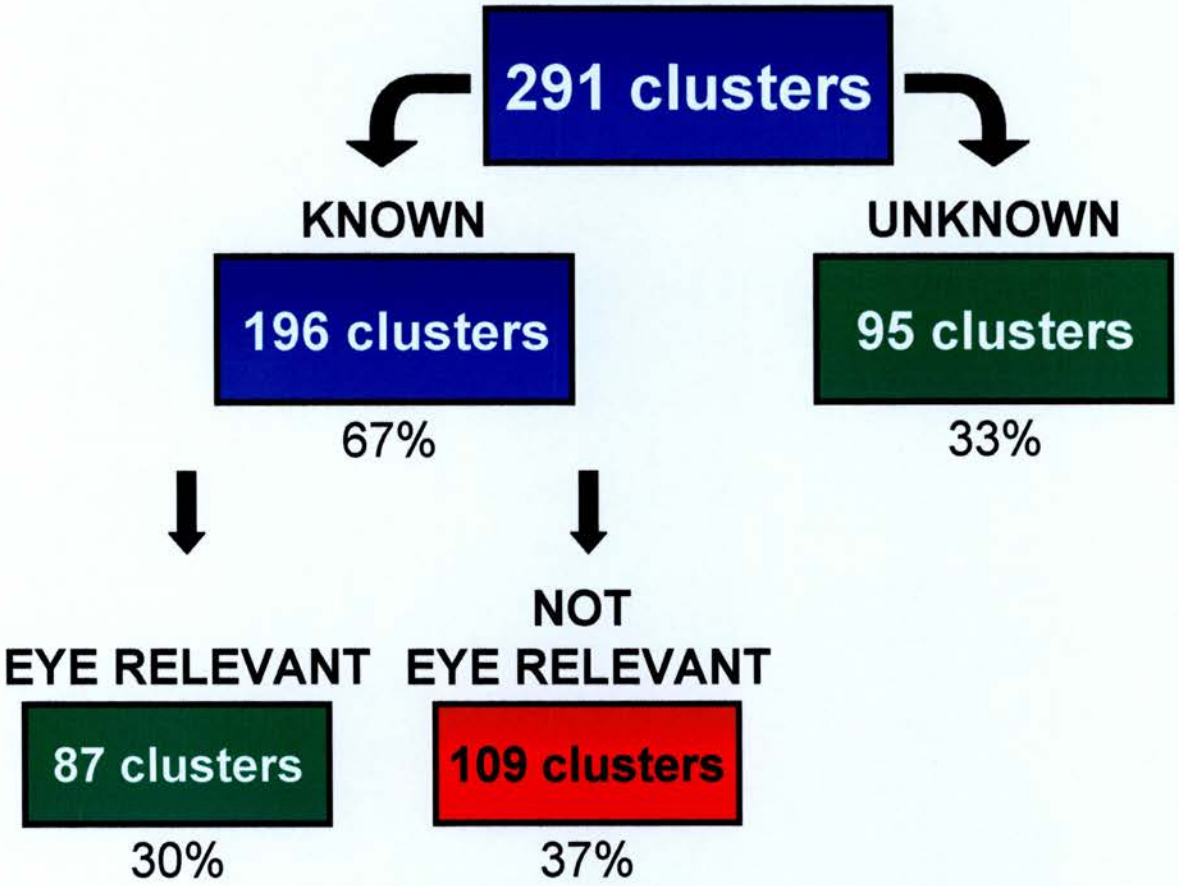


Figure 3.6 Cluster composition of mouse UniGene build 100 DDD results after applying new knowledge from build 120
The total number of clusters with significantly more expression in the eye pool that were retained intact or had sequences adopted by other clusters was 291. The number of clusters is shown in the boxes and the percentage of the total for each category is given underneath. Known clusters are those assigned with a gene name. Eye relevant clusters are those that have been previously reported to have a specific eye expression or function.

3.4 Mouse Digital Differential Display: build 120

3.4.1 Composition of mouse build 120

The third and final mouse DDD was carried out on mouse build 120 in April 2003. The sequence composition of this build is shown (Table 3.1). Since mouse build 100, there has been an introduction of high-throughput cDNA sequences (HTCs). These sequences are longer than ESTs but may include UTRs and introns (Benson et al., 2003). The vast majority of sequences in mouse build 120 are either 5' or 3' ESTs, with these two types have an almost equal contribution (45% and 46% respectively) in contrast to the domination by 3' ESTs in the previous mouse builds. Unclassified ESTs provide a greater proportion (6%) than in the preceding builds, with the minority classes of mRNA and HTCs (1% and 2% respectively) making up the difference.

There are 90,444 clusters in this build, 16,177 (18%) with at least one mRNA sequence, 30,021 (33%) containing at least one HTC sequence, 89,165 (99%) including at least one EST and 15,063 (17%) incorporating both mRNA and EST sequences. The size of clusters ranged from one to more than 8,193 sequences. The cluster distribution amongst size classes is shown (Figure 3.7). The only classes with a lower frequency than in build 100 are the singletons (28,139; 31%) and the 33 to 64 class (3,892; 4%). Again, the general trend of decreasing number of clusters with cluster size applies.

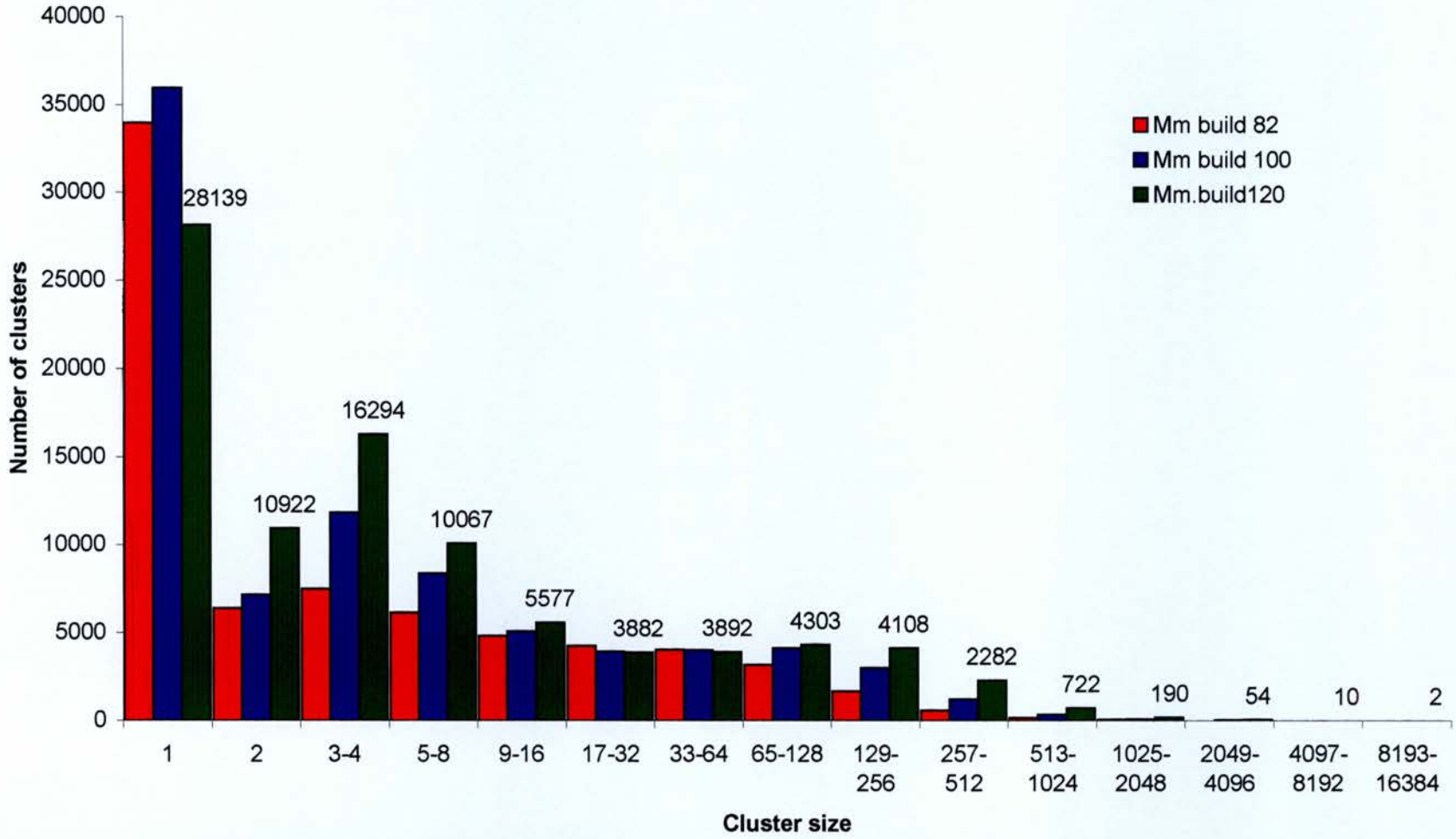


Figure 3.7 Cluster size distribution in mouse UniGene build 120
Cluster size classes are in terms of the number of sequences contained in them and the number of clusters that fall into each class are shown with the relevant number above each bar (green). The distributiosn of mouse UniGene builds 82 (red) and 100 (blue) are also shown.

3.4.2 Library pool construction for DDD on mouse build 120

The eye pool constructed for the DDD on mouse build 120 consisted of seven libraries (E12 eyeball; E15, 16, 17, 18 eye; neonate eyeball; adult eyeball; neural retina; two adult retina) with 78,722 ESTs in total. The target pool comprised 201 libraries containing 1,603,455 ESTs.

3.4.3 Results analysis for DDD on mouse build 120

The DDD results were analysed in the same manner as those of the previous two mouse DDDs, the outcome of which is shown (Figure 3.8). There are a greater number of clusters with significantly more representation in the eye pool than in the previous two DDDs and this has led to a larger number of both known and unknown clusters. The percentage of genes that are known (58%) is lower than the retrospective analyses of the previous DDDs (64% and 67%). The percentages of eye relevant clusters and clusters with no evidence of eye relevancy are both 29%, with an almost equal number of clusters (110 and 107 respectively).

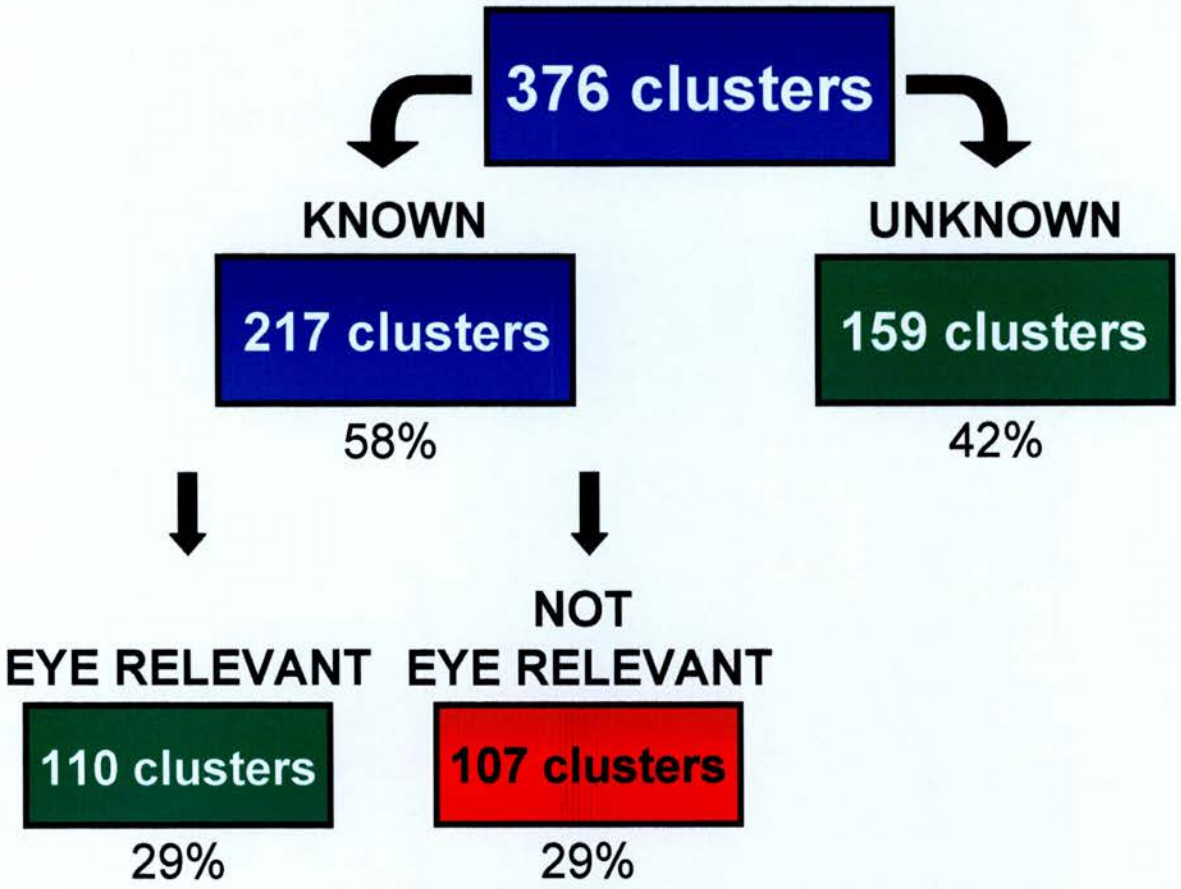


Figure 3.8 Cluster composition of mouse UniGene build 120 DDD results
The total number of clusters with significantly more expression in the eye pool was 376. The number of clusters is shown in the boxes and the percentage of the total for each category is given underneath. Known clusters are those assigned with a gene name. Eye relevant clusters are those that have been previously reported to have a specific eye expression or function.

3.5 Human Digital Differential Display: build 160

3.5.1 Composition of human build 160

The human DDD was executed on human build 160 in April 2003. The numbers of sequence types contributing to this build is shown (Table 3.1). Unlike the situation in the mouse, where 3' ESTs were either the dominant class or equally represented with 5' ESTs (mouse build 120), the largest class of sequence in the human build is 5' EST reads (46%). Three prime ESTs still comprise a significant proportion of the total (35%). There is a larger percentage of unclassified ESTs in the human build (17%) than in mouse build 120 (6%).

The number of clusters present in human build 160 is 111,064, with 27,887 (25%) possessing at least one mRNA sequence, 5,999 (5%) including at least one HTC sequence, 109,660 (99%) containing at least one EST and 26,547 (24%) incorporating both mRNAs and ESTs. Cluster size ranges from one to more than 16,385 sequences. The distribution of clusters amongst size classes is shown (Figure 3.9). As with the mouse builds, a lower number of clusters fall into the larger cluster size categories, although classes 65-128 and 129-256 seem to have more occurrences than the preceding 33-64 sequences category. The singletons account for 36% of the total.

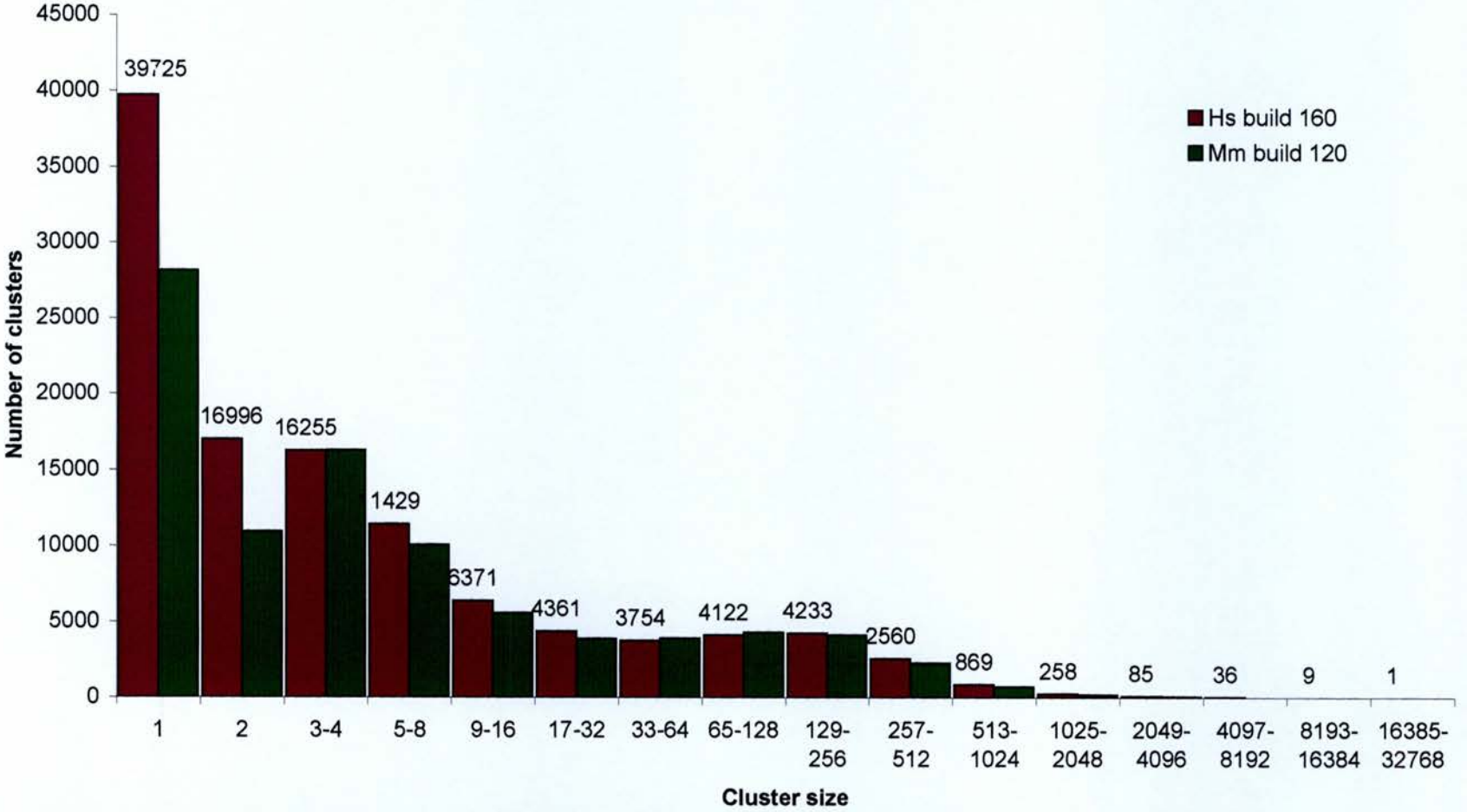


Figure 3.9 Cluster size distribution in human UniGene build 160
Cluster size classes are in terms of the number of sequences contained in them and the number of clusters that fall into each class are shown with the relevant number above each bar (purple). The distribution of mouse UniGene build 120 is also shown (green).

3.5.2 Library pool construction for DDD on human build 160

The eye pool constructed for the human DDD comprised 26 libraries (iris; retinal pigment epithelium (RPE); two foveal and macular retina; two foetal eye and adult lens, eye anterior segment, optic nerve, retina, foveal and macular retina, RPE and choroid; two eye anterior segment; two optic nerve; three lens; three RPE and choroid; four foetal eye; six retina) containing a total of 105,143 ESTs. The target pool consisted of 120 libraries, with 713,026 ESTs, chosen in the same way as the mouse target libraries.

3.5.3 Results analysis of human build 160 DDD

Analysis of the DDD results is shown (Figure 3.10). The total number of clusters showing preferential expression in the eye pool compared to the target pool was 299, with 197 (66%) known. This is similar to the proportion in the first two mouse builds, after applying new knowledge from the current mouse build, but more than the proportion from the results from the current mouse build itself. The major difference between human and mouse in terms of these categories is the lower proportion of eye relevant clusters (27%) to those clusters that have not been found to have any eye relevance (39%).

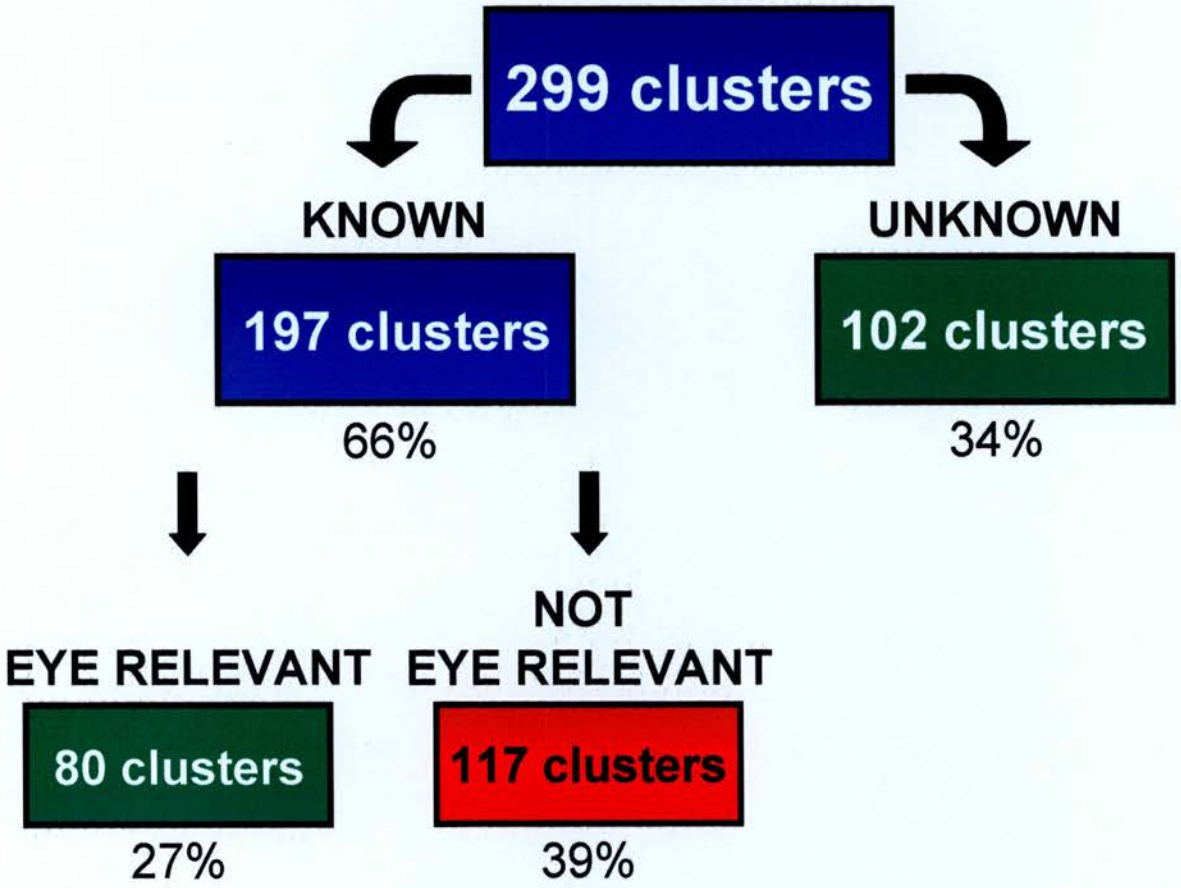


Figure 3.10 Cluster composition of human UniGene build 160 DDD results
The total number of clusters with significantly more expression in the eye pool was 299. The number of clusters is shown in the boxes and the percentage of the total for each category is given underneath. Known clusters are those assigned with a gene name. Eye relevant clusters are those that have been previously reported to have a specific eye expression or function.

3.5.4 Comparison of known genes from human and mouse build DDDs

A table listing the known genes that have significantly greater expression in eye than in non-eye pools is provided (Appendix A). Their gene symbol and name are given according to the accepted nomenclature, verified by MGI or HGNC, with the exception of a few cases where this information could not be found. Their eye relevance and their build identity, that is which build or builds they appeared in, are also indicated. In some cases, the genes have been recorded as having two occurrences in a particular build or one occurrence in a build and another in the new analysis of a build. In these cases, the gene has been represented by more than one cluster in that build. These clusters will have subsequently merged.

Of the total 414 genes, 145 (35%) are eye relevant, 268 are not known to be eye relevant and one gene (mouse *Osbp2* or human *OSBP2*) is eye relevant in the mouse but not known to be in humans. It was removed from further comparisons between eye relevant genes and those with unknown eye relevance, as it is unique in falling in both classes.

To further elucidate any patterns in build identity that may be related to finding genes that are preferentially expressed in the eye and have an eye relevant function, the orthologous clusters were analysed in isolation. Of the 68 genes that are orthologous, 49 (72%) are eye relevant, 18 (26%) have unknown eye relevance and the remaining gene is *Osbp2* or *OSBP2*, the gene that is relevant in mouse but not known to be in human.

It is possible that the effectiveness of DDD producing eye relevant genes differs from build to build or between mouse and human. In order to assess this, the proportions of eye relevant genes and those with unknown eye relevance amongst the known set have been compared between builds, both for the full set and also for the orthologous subset, using histograms (Figure 3.11). There appears to be no significant difference in proportion of eye relevant genes across the mouse builds in the full set of known

genes, although the human percentage appears to be slightly lower. Amongst the orthologous genes, there is a lower percentage of eye relevant genes in the mouse build 82 DDD results compared to other builds, although the number of genes deriving from this build are low and this is unlikely to be significant. Disregarding this build, the histograms confirm that a higher percentage of the orthologous genes are eye relevant compared with that of the full set of known genes.

The eye relevant orthologous genes in the list encompass different types of gene products, such as structural proteins (crystallins), enzymes (phosphodiesterases) and transcription factors (*Rax*; retina and anterior neural fold homeobox). There are eye relevant orthologous genes that have a paralogue in the results that is also eye relevant, for example there are many paralogous crystallin genes. However, there are also a few crystallin genes, structural lens proteins, that do not have orthologues in the DDD results such as *Cryge* and *CRYAB*, which are present in the mouse and human results respectively, but not vice versa. In addition, there are examples in the results of eye relevant orthologues having paralogues that are not known to be eye relevant. For example, *Fscn1* and *FSCN1*, actin bundling proteins, are eye relevant while the mouse paralogue in the results, *Fscn2*, is not known to have any eye relevance. Conversely, there are orthologous genes with unknown eye relevance in the results, such as *Gnb3* and *GNB3*, signal transducer subunits, that have eye relevant paralogues, in this case *Gnb1*. *Eno2* and *ENO2*, genes encoding enzymes involved in glycolysis, are examples of orthologues with no known eye relevance in the results that have a paralogue produced by DDD that also has unknown eye relevance (*Eno1*).

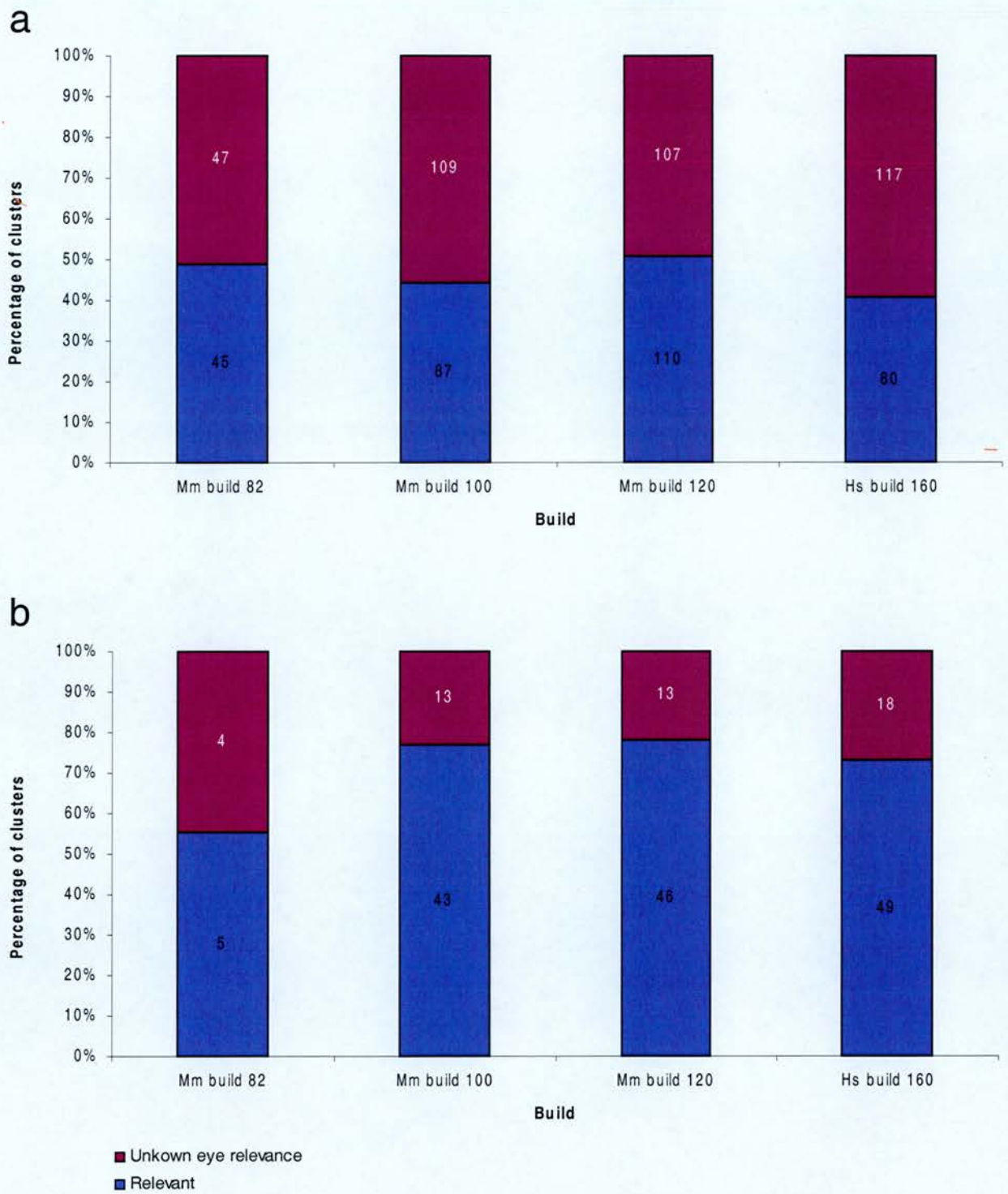


Figure 3.11 Proportions of eye relevant genes amongst subsets of known DDD genes
The percentages of eye relevant genes (blue) amongst all known DDD clusters are approximately equal across all mouse builds, with a slightly lower proportion in the human build (a). Amongst orthologous genes found in both mouse and human DDD clusters, mouse build 82 has a lower percentage of eye relevant genes, although the number of clusters in this category are low (b). A larger percentage of orthologous genes are eye relevant than in the entire known gene set.

3.6 Rat Digital Differential Display: build 119

3.6.1 Composition of rat build 119

The only rat DDD completed was based on rat build 119 in June 2003. The numbers of sequence types contributing to this build is shown (Table 3.1). Unlike the recent mouse and human builds, the majority of sequences in the rat build are 3' ESTs (62%), with the 5' ESTs the second most prevalent sequence type (34%) and the other types of sequence making up the remaining 4%. This is similar to the sequence composition seen in the earlier mouse builds (82 and 100).

There are 54,012 clusters in the rat build, lower than the number in the mouse build (90,444). This is probably a consequence of reduced saturation of the rat transcriptome than of the mouse in the respective UniGene builds. Of the total number of clusters, 5,424 (10%) contain at least one mRNA, with a much lower proportion of the rat build clusters (less than 0.01%) than the recent mouse and human build clusters containing at least one HTC. There were 53,395 (99%) clusters with at least one EST and 4,807 (9%) clusters containing mRNAs and ESTs. These proportions are the same as those seen in the first mouse build (82). Rat cluster sizes range from one to more than 2,049 and their distribution amongst size classes is shown (Figure 3.12). As with mouse and human build compositions, a decrease in the number of clusters corresponds with an increase in their size, with singletons comprising 17,568 (33%) of the clusters.

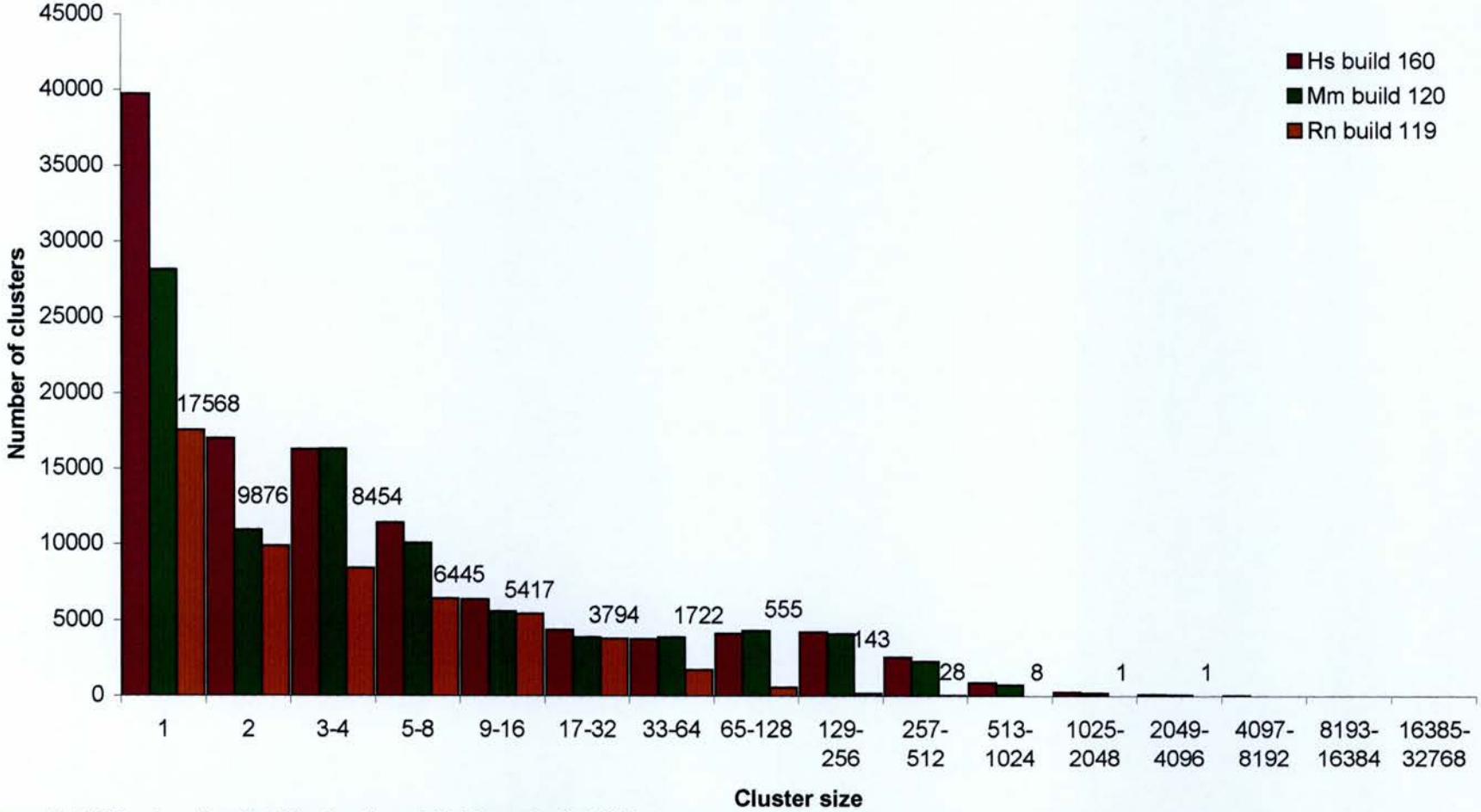


Figure 3.12 Cluster size distribution in rat UniGene build 119

Cluster size classes are in terms of the number of sequences contained in them and the number of clusters that fall into each class are shown with the relevant number above each bar (brown). The distributions of UniGene builds human 160 (purple) and mouse 120 (green) are also shown.

3.6.2 Library pool construction for DDD on rat build 119

There were two eye libraries available for the eye pool in the DDD analysis, both corneal, containing 2,512 ESTs between them. The target pool comprised 17 libraries with 67,587 ESTs.

3.5.3 Results analysis of rat build 119 DDD

The analysis of the rat DDD results is shown (Figure 3.13). Of the 46 clusters that were represented significantly more in the eye pool than the target pool, 26 (57%) were known genes. However, there are no eye relevant genes amongst the known genes. A list of these known genes is provided (Appendix B).

None of these genes have orthologues in either the mouse or human build DDD results. Many of the genes are related to the protective (*C3*; complement component 3) and structural (keratins) function of the cornea.

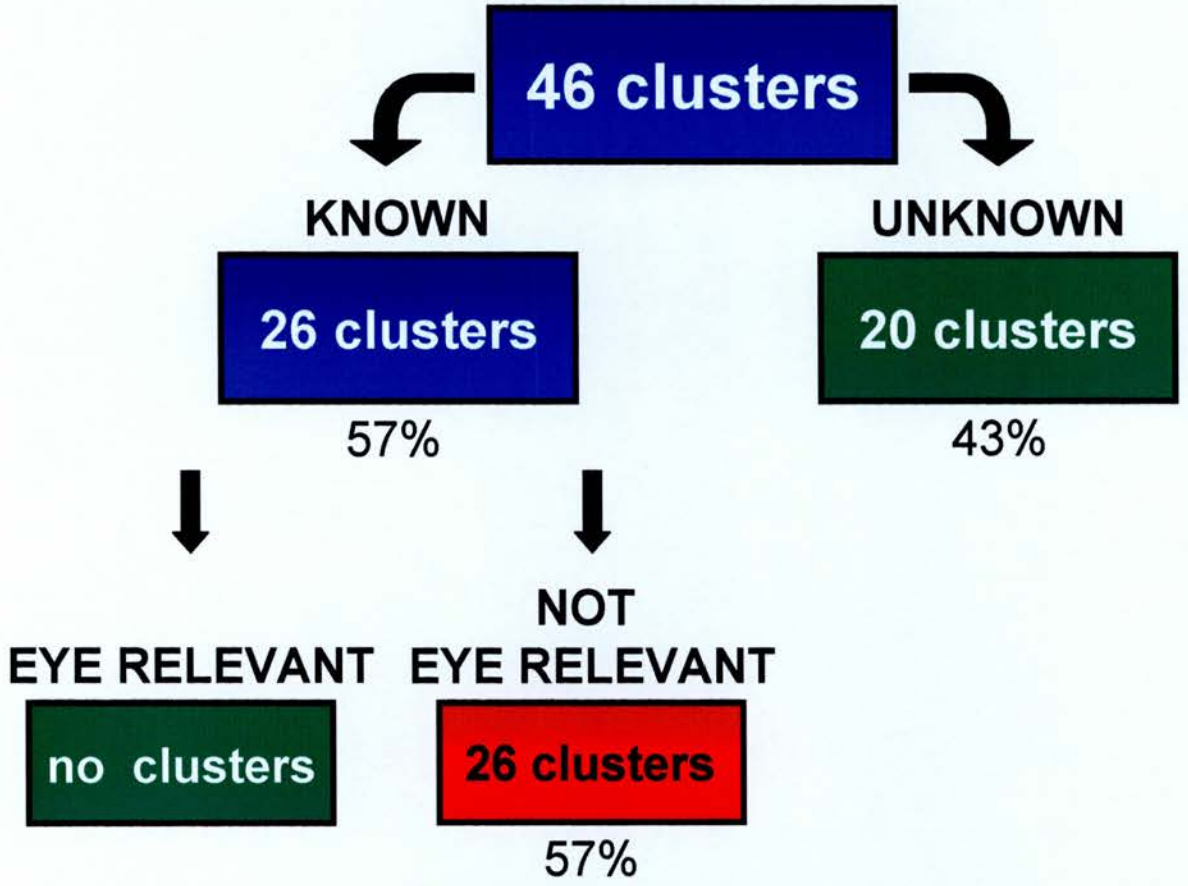


Figure 3.13 Cluster composition of rat UniGene build 119 DDD results
The total number of clusters with significantly more expression in the eye pool was 46. The number of clusters is shown in the boxes and the percentage of the total for each category is given underneath. Known clusters are those assigned with a gene name. Eye relevant clusters are those that have been previously reported to have a specific eye expression or function.

3.7 Producing a set of clusters for further study

The results derived from the mouse build 100 DDD were used as a basis for subsequent investigation. There were 138 unknown clusters produced from the DDD results of this build, each potentially representing a novel gene expressed preferentially in the eye. There was a necessity to refine and reduce the number of clusters to consider by discarding clusters that may not have represented real genes and improve the likelihood that these clusters would be specifically or preferentially expressed in the eye. In lieu of this, only clusters meeting certain criteria were studied further. The first three of the following four criteria were based on the profiles of known eye relevant genes in the mouse build 100 results:

1. Clusters have a six-fold difference between eye pool and target pool representation.
2. Clusters have a maximum of 0.006% representation in the target pool.
3. Clusters have a minimum of 0.01% representation in the eye pool.
4. Clusters contain a sequence annotated as full-length.

Of the 138 unknown genes, only 31 clusters met all four criteria and were therefore used for further study. These clusters are summarised (Table 3.2), with their current UniGene identifier, their representative sequence, which allows tracking of cluster sequences should the cluster itself be retired, and a brief description based on information discovered by carrying out BLAST on mouse Ensembl. These clusters encompass a range of gene types, with possible transcription factors, structural proteins, enzymes and signalling proteins. Through this process, three of these clusters were determined as known genes (Mm.134516, *Otx2*; Mm.89477, *Crygf*; Mm.95707, *Aipl1*). These were subsequently confirmed by UniGene, with a further seven clusters discovered as known genes (Mm.121647, *Rdh12*; Mm.44683, *Gabt4*; Mm.29145, *D15Ertd806e*; Mm.29846, *Ndr4*; Mm.34901, *Grtp1*; Mm.37819, *D5Bwg0860e*; Mm.59151, *Guca1b*).

UniGene cluster	Current identifier	Representative sequence	Description
Mm.103712	Mm.103712	AK014383	coils
Mm.121647	Mm.121647	BC016204	*Rdh12*
Mm.134224	Mm.134224	AK020935	-
Mm.134516	Mm.134516	BC017609	*Otx2*
Mm.137178	Mm.44683	AK021402	*Gabt4*
Mm.150838	Mm.150838	AK021382	-
Mm.151918	Mm.151918	AK018494	MCM domain
Mm.159861	Mm.159861	AK020813	-
Mm.195923	Mm.89477	AK013953	*Crygf*
Mm.200734	Mm.29145	AK004048	*D15Ertd806e*
Mm.213208	Mm.213208	BC017153	bipartite nuclear localisation signal; thioredoxin domain 2
Mm.21485	Mm.21485	BC017540	nucleosome assembly protein domain
Mm.22913	Mm.22913	AK007745	-
Mm.23434	Mm.23434	AK005271	-
Mm.2965	Mm.218574	AK018181	-
Mm.29846	Mm.29846	BC006595	*Ndr4*
Mm.3124	Mm.3124	BC009660	protein kinase binding
Mm.34901	Mm.34901	AK017197	*Grtp1*
Mm.35511	Mm.35511	AK013554	homeobox and transmembrane domains
Mm.37819	Mm.37819	AK018184	*D5Bwg0860e*
Mm.38347	Mm.38347	AK005160	BTB/POZ domain
Mm.44242	Mm.44242	AK009408	related to prosaposin
Mm.44292	Mm.44292	AK019578	RNA recognition motif/pre-mRNA splicing
Mm.56565	Mm.268718	AK017115	Zn finger, C2H2 type; KRAB box
Mm.59151	Mm.59151	BC018258	*Guca1b*
Mm.62589	Mm.62589	AK016342	-
Mm.78118	Mm.78118	AK014852	beta-galactosidase/lactase
Mm.82341	Mm.82341	AK014271	rhodopsin-like GPCR/bipartite nuclear localisation signal
Mm.87130	Mm.87130	AK019549	immunoglobulin and bactericidal permeability motifs
Mm.95707	Mm.95707	AF296412	*Aipl1*
Mm.95741	Mm.95741	AK020807	-

Table 3.2 The set of 31 clusters for further study
The cluster identifier and the representative sequence given at the time of mouse build 100 are shown along with a brief description based on Ensembl BLAST.

3.8 Discussion

The ultimate aim for the UniGene database is to produce a set of clusters that represent the transcriptome of a particular organism. The most recent mouse UniGene build (120) has a total of over 90,000 clusters and the human equivalent (build 160) has over 111,000 clusters. These are well over twice the number of the current estimates of protein-coding genes, which range between 22,000 and 30,000 in mouse (Marra et al., 1999; Waterston et al., 2002) and between 23,000 and 40,000 in human (Lander et al., 2001; Venter et al., 2001; Waterston et al., 2002). The estimate for total number of genes, including those not translated, is of course much greater than that for protein-coding genes, and in mouse is estimated to be approximately 63,000 (Carninci et al., 2003). This is still well below the 90,000 clusters in the current UniGene build. Although the sequences are screened for contaminants before build construction, it is probable that many are not removed, as evidenced by the removal of many sequences from UniGene builds over time. However, this cannot be the only cause of the discrepancy between Carninci et al's estimate and the number of clusters in UniGene. Another factor is redundancy, which may be a true redundancy, where more than one cluster represents the same transcript, or it may be that multiple clusters represent the same gene, but each one has sequences that define separate alternative gene transcripts. As part of Lander et al's human genome sequence publication, it was declared that for each human gene on chromosome 22, an average of 2.6 splice variants were detected and for chromosome 19 this average was 3.2 (Lander et al., 2001). With 70% of splice variants predicted to affect coding sequence (Lander et al., 2001) and 40% of human genes being alternatively spliced (Brett et al., 2000), this may have an effect in the UniGene builds in terms of number, as well as on the complexity of the species' proteomes.

The high number of singleton clusters in the UniGene build may be another indication that there is a high level of redundancy in both the mouse and human builds. The alternatives are a low saturation of gene representation by ESTs or that

singleton sequences may be genomic contamination. The number of singletons alone exceeds current gene number estimates in both the mouse and human, making low saturation of the transcriptome unlikely to be a major factor. The percentage of singleton clusters has decreased from build to build, from 47% in mouse build 82 to 40% in mouse build 100 and finally 31% in build 120, similar to the 36% of singletons in the human build. The decrease from build 100 to 120 is accompanied by a decrease in the actual number of singleton clusters, even though the total number of clusters has increased. This decrease in singletons could be symptomatic of a decrease in redundancy or a more effective removal of contaminating sequences. However, as the proportion of singletons in the builds is, at its lowest, over a third of the total, the implication is that many singletons represent real mRNAs. Estimates of polyadenylation signal frequency in a recent large-scale sequencing project indicated that, in their data, this proportion was a half (Carninci et al., 2003). As the clusters chosen from DDD results for further analysis were full length, this effect should not influence this subset.

The number of clusters showing significantly different levels of expression between the eye and target pools has increased from the earlier to the later mouse builds, probably due to an increase in the total number of clusters from build to build. When applying current build knowledge to all mouse builds, the proportion of known genes is higher than the unknown genes. This was not true of the original analysis of the build 82 DDD results, but a number of clusters have been removed since, presumably due to contaminating sequences. An additional factor is gene discovery, with the number, as well as the proportion, of known genes increasing. The proportion of eye relevant genes is consistently between 29% and 31% when applying new knowledge to all mouse builds. However, the proportion of genes that are not eye relevant lies between 29% and 37% which is higher than the eye relevant proportion in the first two builds but equal to the recent mouse build. The proportion of known genes from DDD results that are eye relevant, given the current knowledge of the genes involved, is between 45% and 50%. If this is applied to the mouse build 100 DDD unknown genes, 72 to 80 of the 159 clusters will be eye relevant genes. However, this is likely to be an overestimate of the true number. There will be clusters in the

DDD results containing ESTs not representative of a true mouse gene, either due to genomic DNA or the products of illegitimate transcription, for example, promoter activity from *Alu* sequences, which remain untranslated. In addition, redundancy is more likely to occur in the unknown cluster set as there are clusters without full-length sequences and, as a consequence, a greater chance that ESTs deriving from the same transcriptional unit will not be clustered.

For the human build DDD results, the proportion of eye relevant genes in the known gene set is slightly lower than those in the recent mouse build DDDs, but there is also a lower number of clusters (80 compared to 110 clusters in the recent mouse build 120 DDD results). This may reflect a difference in the libraries available for use in the mouse and human DDDs. It is also possible that this may relate to the differences in the way these two organisms are studied rather than an intrinsic difference in the phenotypic consequences between the two genomes. A large proportion of mouse phenotypes derive from induced mutation and can be studied more invasively and in larger numbers. However, since human mutations are existing rather than induced, they are present in smaller numbers and the phenotypes are not as fully described and studied as those of mice.

The presence of orthologous genes, found in the mouse and human DDD results, gives greater confidence in the eye relevancy of that gene, shown by the higher proportion of eye relevant genes amongst the orthologues than across the whole list. However, this still leaves 19 genes that are not known to be eye relevant in the mouse but have orthologues amongst the human results. This may be part of the error in the DDD process, or these may be genes that have an as yet undiscovered eye relevant function. The same is true for the paralogues that are present in the DDD results.

One of the orthologues that is not known to be eye relevant in the mouse is *Osbp2* (oxysterol binding protein 2) and is unique amongst the orthologues in having a splice variant with reported high and specific expression in human retina (Moreira et al., 2001), but no known eye relevant role in mice. This difference could either be

due to a real difference between mouse and human in gene expression, or it may be an indication that the eye expression of *Osbp2* has not yet been revealed in mice. The latter is likely to be true, since mouse studies have commented on the possible eye relevant role of *Osbp2* (Anniss et al., 2002), but have not as yet demonstrated a link.

The genes deriving from DDD have been described as eye relevant or not known to be eye relevant. It is almost impossible to declare any gene to be known not to have a particular function, as new functions for genes are being continuously described, thus the known functions of a gene cannot be presumed to be the only ones. It is therefore possible that some of the genes not known to be eye relevant may have an undescribed optic function. This is more likely when orthologous genes have been produced in the results. The orthologous genes *Gnb3* and *GNB3* (guanine nucleotide binding protein, beta 3), as well as *Gnb5* and *GNB5* (guanine nucleotide binding protein 5), are not known to be eye relevant. However, their paralogue *Gnb1* (guanine nucleotide binding protein, beta 1), also identified by DDD, has an eye specific function as part of the beta subunit of retinal transducin (Danciger et al., 1990). *Gnb3* and *GNB3* form the beta subunit of transducins, however *GNB3* has been found to be ubiquitously expressed (Levine et al., 1990). *Gnb5* is predominantly expressed in brain (Watson et al., 1994) while *GNB5* is expressed in brain, pancreas, kidney and heart (Jones et al., 1998). Another pair of orthologous genes in the results that are not known to be eye relevant are *Eno2* and *ENO2* (enolase 2, gamma neuronal), with their non-relevant paralogue, *Eno1* (enolase 1, alpha non-neuron), also produced in the results. Enolases are glycolytic enzymes, with *Eno2* specific to neurons and *Eno1* ubiquitously expressed (Lebioda and Stec, 1991). Of possible relevance is the fact that tau crystallin, a type found in reptiles, birds and lamprey, has been identified as an enolase, with a much higher abundance in the lens than required for its enzymatic activity in other tissues (Wistow et al., 1988). It is possible that some or all of these orthologous genes with unknown eye relevance may have an eye specific alternative transcript, as with *OSBP2*, or an undiscovered eye specific function.

A number of the known genes with unkown eye relevance in the results have been implied, through differential expression but not by the criteria required for eye relevance described earlier (section 3.2.3), to have an eye relevant role in previous studies. In a study to identify genes differentially expressed in cow RPE and retina, the bovine homologue of *Wdr6* (WD repeat domain 6), with unknown function, was found to be upregulated (Sharma et al., 2002) and the canine homologues of *Dnm* (dynamin, involved in endocytosis) and *Ubc* (ubiquitin C, associated with protein degradation) were found to be differentially expressed in retina, with elevated expression (Lin and Sargan, 2001). *CPE* (carboxypeptidase E, involved in hormone synthesis), and *C4A* (complement component 4A, associated with immunological defence) were identified as genes upregulated in the human ciliary epithelium (Coca-Prados et al., 1999). Other genes displaying differential expression in eye domains have also been found to have other eye relevant connections. For example, *TF* (transferrin, iron transporter) and *ALDOC* (aldolase C, fructose-biphosphate, a glycolytic enzyme) have been extracted from an enriched human retina library (Sinha et al., 2000) and been found to be photoreceptor specific in a separate study (Swanson et al., 1997). Further to these studies, *TF* has been found to be upregulated in the retinal periphery compared to the macula (Sharon et al., 2002) and *ALDOC*, as well as *LOX* (lysyl oxidase, a crosslinking enzyme) and *SLPI* (secretory leukocyte protease inhibitor), was discovered to be upregulated in corneal epithelium derived from individuals with the human corneal disease, keratoconus, when compared to normal cornea. *IGFBP5* (insulin-like growth factor-binding protein 5) has been extracted from an enriched human retinal library (den Hollander et al., 1999) and the rat homologue has also been found, together with *Aldo1* (aldolase 1), to be upregulated in the retina of a diabetes-inducible strain after chemical administration, compared to pre-treatment levels (Joussen et al., 2001). Four genes from the known gene list (*CLU*, clusterin; *CD74*, CD74 antigen; *EEF1A1*, eukaryotic translation elongation factor 1, alpha-1;*SERPINA3*, a protease inhibitor) are upregulated in the peripheral retina of an older eye compared to that of a younger eye (Sharon et al., 2002). In the same study, *Ndr4* (N-myc downstream regulated 4) shows higher expression levels in retinal periphery than macula, while *Nefl* (neurofilament, light polypeptide) and *Tuba1* (tubulin, alpha 1) have higher expression in the macula.

Tuba1 is also upregulated in human trabecular meshwork, through which the outflow of aqueous humour passes, in response to an induction of raised intraocular pressure, a situation that is associated with glaucoma, and *Scg2* (secretogranin II), also shows this trend (Gonzalez et al., 2000b). Thirdly, *Tuba1* has also been found to be more highly expressed in $Crx^{-/-}$ mice (mice without the expression of this transcription factor in developing photoreceptors) than in normal mice, along with *Tuba2*, *Nedd4* (neural precursor cell expressed, developmentally down-regulated gene 4), *Rpl13* (ribosomal protein L13) and *Rpl7a* (ribosomal protein L7a) (Blackshaw et al., 2001). In this analysis, *Sast* (syntrophin associated serine/threonine kinase), *Gpi1* (glucose phosphate isomerase), *Eno2* and *Tubb5* (tubulin beta 5) are shown to be downregulated in $crx^{-/-}$ mice. Along with *Ddr1* (discoidin domain receptor family member 1, a receptor tyrosine kinase), *Tubb5* also has lower expression in the transgenic mouse lenses overexpressing PAX6 than in haplosufficient PAX6 mouse lenses (Chauhan et al., 2002). *KRT5* (keratin 5) has been seen to be downregulated in the corneal endothelium of individuals with Fuch's dystrophy, a progressive corneal disorder, compared to control corneal epithelium (Gottsch et al., 2003) and, finally, *Crabp1* (cellular retinoic acid binding protein 1) has elevated expression in the ventral, compared to the dorsal, side of the developing mouse retina (Diaz et al., 2003). Taken together, these studies have shown differential expression in eye tissue, or in certain domains of normal, diseased or experimentally challenged eye tissue, for 31 of the 267 non-relevant genes deriving from DDD, with five of these mentioned in more than one study. This further indicates that there are genes that do not display an obvious eye relevant role but may nevertheless be differentially expressed and thus important in eye development and function.

It was expected that there would be a number of genes that have eye relevance, in terms of association with an eye disease, that would not appear in the list. These include *Rpgr*, discussed earlier (section 1.5.4), which is ubiquitously expressed. As hypothesised, this gene is not produced in the mouse DDD results, as is the case with its human homologue in the human DDD results. In contrast, the specifically expressed *Rho* and preferentially expressed *Pax6* are both produced in the mouse results along with their human counterparts in the human DDD results. The alpha

crystallin genes, *Cryaa* and *Cryab*, also discussed earlier (section 1.5.3), differed; *Cryaa* was specifically expressed, while *Cryab* was more widely expressed, but still with significant expression levels in the eye compared with other tissues. *Cryaa* and *CRYAA* were present in the mouse and human DDD results respectively, as would be expected. However, although *CRYAB* was present in the human DDD results, *Cryab* was not produced in those of the mouse. This may represent a real difference in the expression between mouse and human. Alternatively, it may be that the larger number of transcripts in the human has produced the expected, and perhaps more accurate, result of preferential eye expression.

The rat DDD was initiated with the expectation of producing more eye relevant genes, some of which may have been orthologues of genes from the mouse and human DDD results. However, the lower number of gene sequences and ESTs available for the rat than for the current mouse build has led to a sequence composition comparable to the earliest mouse build (82) and there is a lower number of libraries and a skew in their composition. With no eye relevant genes produced, this was an unsuccessful attempt to confirm mouse genes by producing their rat orthologues. It demonstrates the effect that the eye libraries have, and it may be that the composition of the eye pools in the mouse and human builds are skewed in favour of certain eye domains. In terms of the number of libraries, this may have been the case in mouse build 100, when four of the seven libraries were retinal, and in human build 160, which had 15 of 22 specific eye domain libraries (26 libraries in total) having a retinal component. However, it is the number of sequences contributed by the different eye domains that will influence the results and there is no way of judging what the 'correct' representation of each of these domains should be, nor is this a factor that could, or may need to be, easily controlled. Indeed, any retinal favour in the pool that there may be is not excluding genes important in structure and function of other eye domains, such as crystallins (lens) and keratocan (cornea). This does not preclude the possibility that genes expressed at lower levels in these tissues may not be represented in the eye pool sufficiently to be discovered in this way. As well as the organ or tissue domain composition of the library pools, other features of

the libraries' identities could have an influence on the results, such as sex, which has been previously shown to have a differential effect (Sun et al., 2000).

Ten of the 31 clusters that were chosen for further study have turned out to represent known genes. Three of these were determined to be such at the time of DDD initiation and were later annotated as genes in the UniGene build. The first of these, *Otx2* (orthodenticle homologue 2), is a homeodomain protein related to the *orthodenticle* gene in *Drosophila* that is important in head development (Simeone et al., 1992) and essential for eye morphogenesis, with $Otx2^{-/-}$ mice having severe eye defects (Matsuo et al., 1995). The second known gene, *Crygf* (crystallin gamma F) is one of many crystallin genes that encode structural proteins in the lens (Goring et al., 1992). However, unlike the other crystallins, gamma crystallins are expressed in the elongating central nuclear fibres of the adult lens as well as the secondary fibre cells (Van Leen et al., 1987). The $Crygf^{Rop}$ mouse, recovered from a mutagenesis screen, has a dominant cataract phenotype (Graw et al., 2002). Of these three known genes, *Crygf* is the only one not to have a human homologue in the human DDD results. The third known gene is *Aipl1* (aryl hydrocarbon receptor-interacting protein like-1), the human homologue of which is *AIPL1*. Mutations within *AIPL1* have been found to be associated with a severe, early-onset form of LCA (Leber congenital amaurosis), an autosomal recessive retinal degeneration disorder (Sohocki et al., 2000). Although the function of AIPL1 is not known, in the adult human retina it has been found to be expressed specifically in the rod photoreceptors (van der Spuy et al., 2002).

In addition to these three clusters that represented genes known at the time, an additional seven clusters represent genes that have been annotated as known genes in the current mouse UniGene build (120), with six of these seven representing genes that have been described recently. *Rdh12* (retinal dehydrogenase 12) was first isolated from human retina. Retinal sections showed it to be specifically expressed in monkey and mouse photoreceptors, although it is expressed in other tissues (Haeseleer et al., 2002). *Gabt4* (gamma-aminobutyric acid (GABA-A) transporter 4) encodes one of four transporters for the GABA neurotransmitter and is specific to

neural tissue in mouse (Liu et al., 1993) and shows particularly strong expression in the rodent spinal cord embryonically (Jursky and Nelson, 1999). *Ndr4* (N-Myc downstream regulated 1) is the fourth member of a family and in humans NDR1, NDR2 and NDR3 are ubiquitously expressed but in contrast NDR4 is specifically expressed in brain and heart (Zhou et al., 2001). Rat NDR4 has been implicated in having a role in neurite outgrowth (Ohki et al., 2002). *Grtp1* (GH regulated TBC protein 1) was identified from a microarray analysis of GH-responsive cells *in vitro* and expression analysis showed it to be expressed highly in testis, with lower levels in intestine, kidney, liver and lung (Lu et al., 2001). *Guca1b* (guanylate cyclase activator 1B) was originally studied as a gene that encoded a ligand of guanylate cyclase-C, an intestinal receptor (Currie et al., 1992) and was found to be highly expressed in the proximal small intestine of mice (Whitaker et al., 1997). However, it is also expressed in the mouse retina, predominantly the rod photoreceptors, as well as in the amacrine and ganglion cells (Howes et al., 1998), where it is thought to activate guanylate cyclase during phototransduction (Mendez et al., 2001). The final two genes were discovered as part of large-scale studies, with *D5Bwg0860e* being mapped as a gene on chromosome 5 by SSCP (Brady et al., 1997) and *D15Ertd806e* was mapped as a gene on chromosome 15 and derived from the two cell stage of an embryo (Ko et al., 2000).

UniGene statistics and DDD results have shown that the most recent builds are best for analysis, with a greater number of clusters but an implied reduction of redundancy. It may therefore be useful to check older DDD results with those from the most recent build. Another effective method to validate genes deriving from a mouse DDD would involve comparison with the human equivalent, as a higher proportion of orthologous genes are eye relevant. These methods were not employed for selecting clusters for more investigation in this study. However, five (*Otx2*, *Crygf*, *Aipl1*, *Rdh12* and *Guca1b*) of the 10 known genes amongst the 31 clusters chosen using other criteria for further investigation are associated with important roles in eye development or function. Thus their discovery serves as a vindication of the DDD method utilised to discover novel genes with eye relevant roles. The next

two chapters describe expression analysis of the novel clusters to confirm preferential expression in the eye experimentally.

CHAPTER 4

Analysis of novel preferentially expressed clusters

4.1 Introduction

In order to make an initial assessment of the success of DDD in extracting novel clusters that are preferentially expressed, further analysis was carried out on the 31 clusters derived from the UniGene build 100 DDD output. This entailed carrying out an RT-PCR screen to select clusters appropriate for further investigation. Subsequently, expression studies were carried out using *in situ* hybridisation during developmental stages as well as in the neonate and adult.

Mouse eye development begins at E7.5, with the lens and cornea deriving from surface ectoderm, neural ectoderm giving rise to the retinal layers, ciliary body and iris and neural crest creating Descemet's membrane, the sclera and the pigment layers in the iris, ciliary body and retina. By E9.5, the optic vesicle has matured to form the optic cup, or secondary optic vesicle, and the optic stalk (Figure 4.1). The presumptive neural and pigment layers of the retina derive from the inner and outer layers of the optic cup respectively and begin to develop at E10, with the pigment layer differentiated by E11. Meanwhile, at approximately E10, the lens placode invaginates to give rise to the lens pit, which eventually buds off to form the circular lens vesicle by E10.5. The choroid has differentiated by E11 and the lens fibre cells begin to form between E11.5 and E12.5, as the posterior cells of the lens vesicle elongate in the vesicle. At E11.5, the corneal epithelium begins to develop from the surface ectoderm that lies adjacent to the lens vesicle. By E12.5 the optic nerve and the choroid, deriving from the optic stalk, have formed, the lens, sclera, cornea and pigment layer are undergoing maturation and the neural layer of the retina is differentiating.

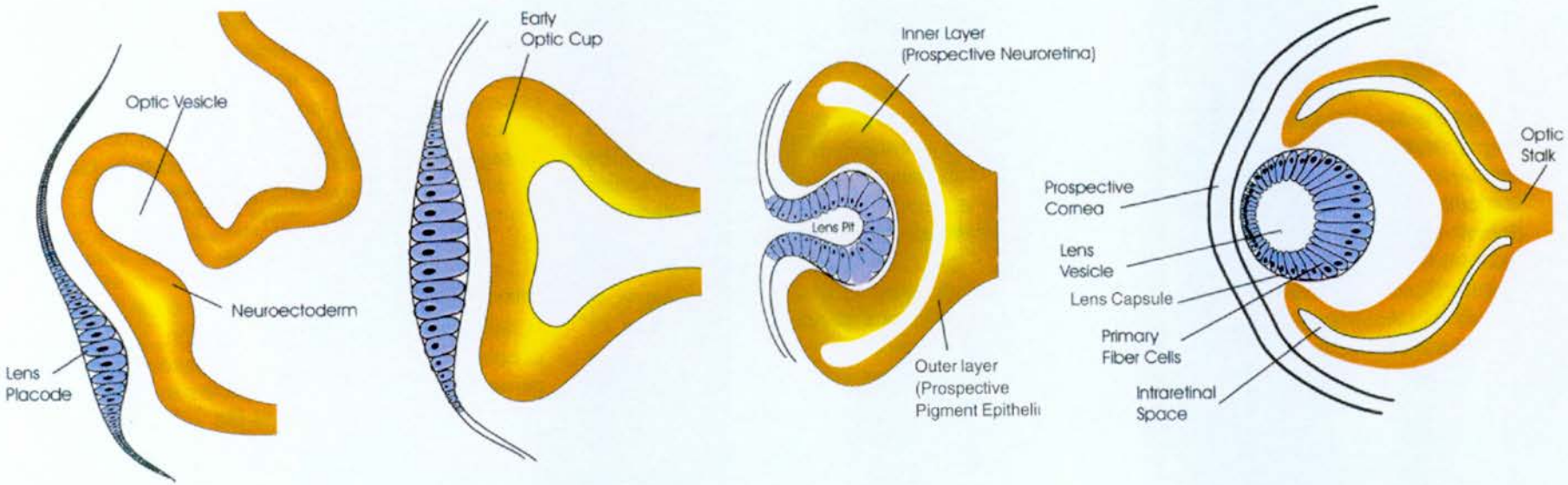


Figure 4.1 Early eye development
As the optic vesicle develops into the optic cup between E9 and E10, the lens placode begins to invaginate. The lens pit develops as the optic cup differentiates into the inner neural retina and the outer presumptive RPE. By E12.5, the future cornea has begun to develop from the surface ectoderm and the lens pit has formed a vesicle, with the fibre cells extending centrally towards the anterior lens cells. The optic stalk has also formed, while the neural retina is differentiating and the RPE undergoes maturation. Adapted from Cvekl et al.

The lens vesicle becomes the lens capsule at about E12.5, when the lens fibre cells have elongated to obliterate the vesicle cavity. After this time, the nuclei of the lens fibres migrate away from the periphery to a central position within the cell and the lens capsule, with the cells becoming more transparent from E17.5 to birth. After birth, the lens enlarges through the genesis of additional secondary lens fibre cells from the equator and is visibly mature at P12 to P14. Meanwhile, the choroidal layer undergoes development, with its capillary network developing from E12.5 and its neural crest-derived pigment cells differentiating at E12.5. The sclera differentiates at E15.5 and the corneal endothelium and Descemet's membrane begin to differentiate at approximately E16, with the cornea undergoing development until after birth. Retinal development also continues after birth. The ciliary body and iris are formed at E16.5 from the peripheral neural retina. Simultaneously, the retinal layers begin to differentiate visibly at approximately E16.5 or E17.5. Initially, there are three layers: the outer nuclear layer, consisting of horizontal cells and the nuclei of photoreceptors; the transient intermediate layer; and the inner nuclear layer, containing supporting cells and ganglion cells, which later form axons to the optic stalk. The cell layers forming the adult mouse retina (ganglion cell layer, inner nuclear layer, outer nuclear layer and the retinal pigment epithelium) are not recognisable until several days after birth.

Bearing in mind the processes taking place at various developmental stages, the expression patterns of clusters expressed preferentially in the eye, along with their bioinformatic investigation, have been used as a basis for hypothesising the possible functions of these clusters.

4.2 Expression analysis using RT-PCR

In order to validate the preferential expression of clusters, RT-PCR was utilised to assess expression in CD1 mouse eyes taken at four stages (E14.5, E16.5, neonate (P1) and adult) as well as in a panel of seven other adult CD1 tissues (brain, heart, kidney, liver, lung, ovary and testis). The primer pair for each cluster was designed to amplify its representative sequence, as given by UniGene, and a *Gapd* primer pair was also used as a control to confirm that RNA used from samples had approximately equal quantities and were not degraded.

On the basis of the results, the clusters were classified as displaying eye specific expression, preferential eye expression (expressed in three or less non-eye tissues) or non-preferential eye expression (ubiquitous expression, no eye expression or undetectable expression). The results of the four eye specific, 11 preferentially eye expressed and 13 non-preferentially expressed clusters are shown (Figures 4.2, 4.3 and 4.4 respectively). Three clusters did not produce detectable expression in these tissues by RT-PCR (Mm.44292, Mm.62589 and Mm.78118). The RT-PCR results of tested clusters are summarised (Table 4.1).

Of the four eye specific clusters, three are known genes (Mm.195923, *Crygf*; Mm.95707, *Aipl1*; Mm.59151, *Guca1b*). The remaining clusters are therefore potentially novel genes, with Mm.150838 showing expression in the eye at every stage but the adult.

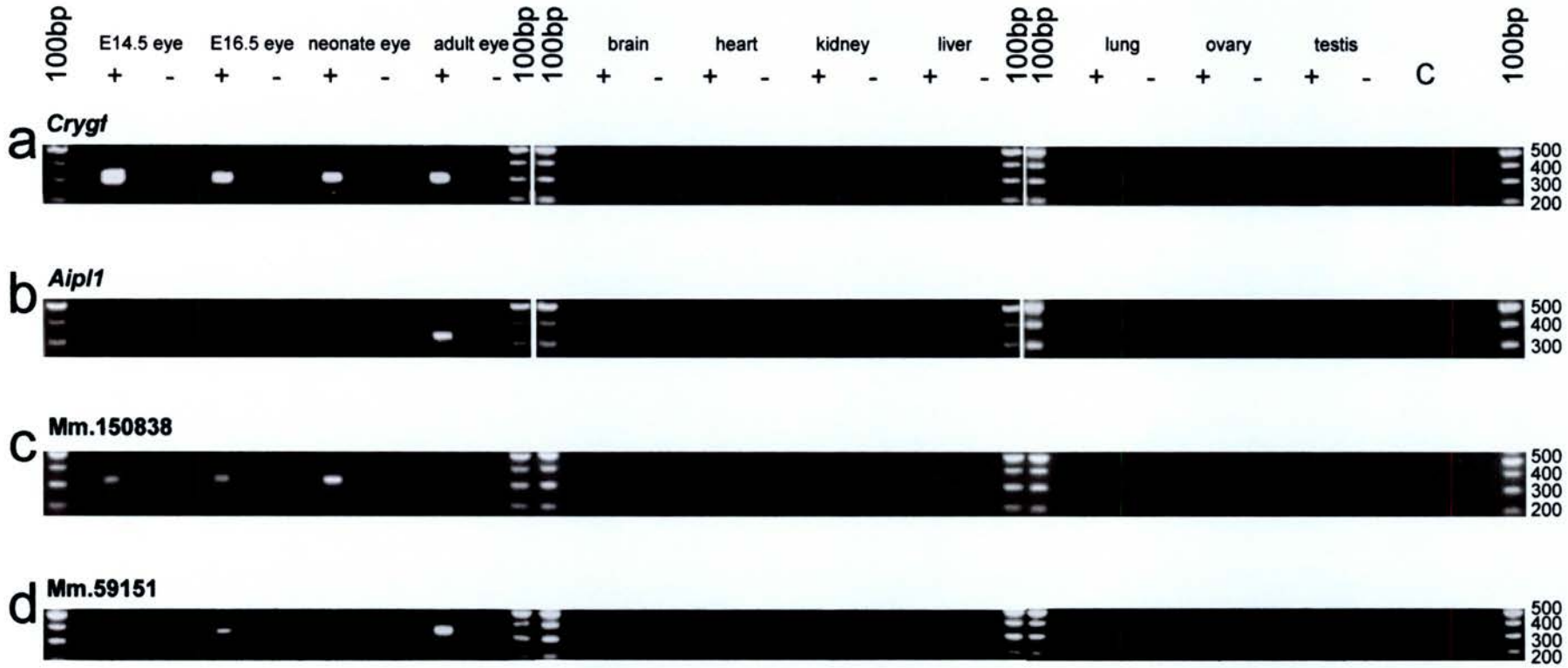


Figure 4.2 RT-PCR panels of clusters showing eye specific expression
RT-PCR expression is eye specific for four clusters, two of which are known genes. Eye specific expression is shown during both eye development and in the adult (a, d), during eye development only (c) and purely in the adult (b). For each tissue, a reaction using a reverse transcription with no enzyme was carried out (-) in addition to the experimental reaction (+). An additional negative control with water is shown (C). All non-eye tissues are adult in origin. The number of base pairs represented by bands in the 100bp marker is shown on the right.

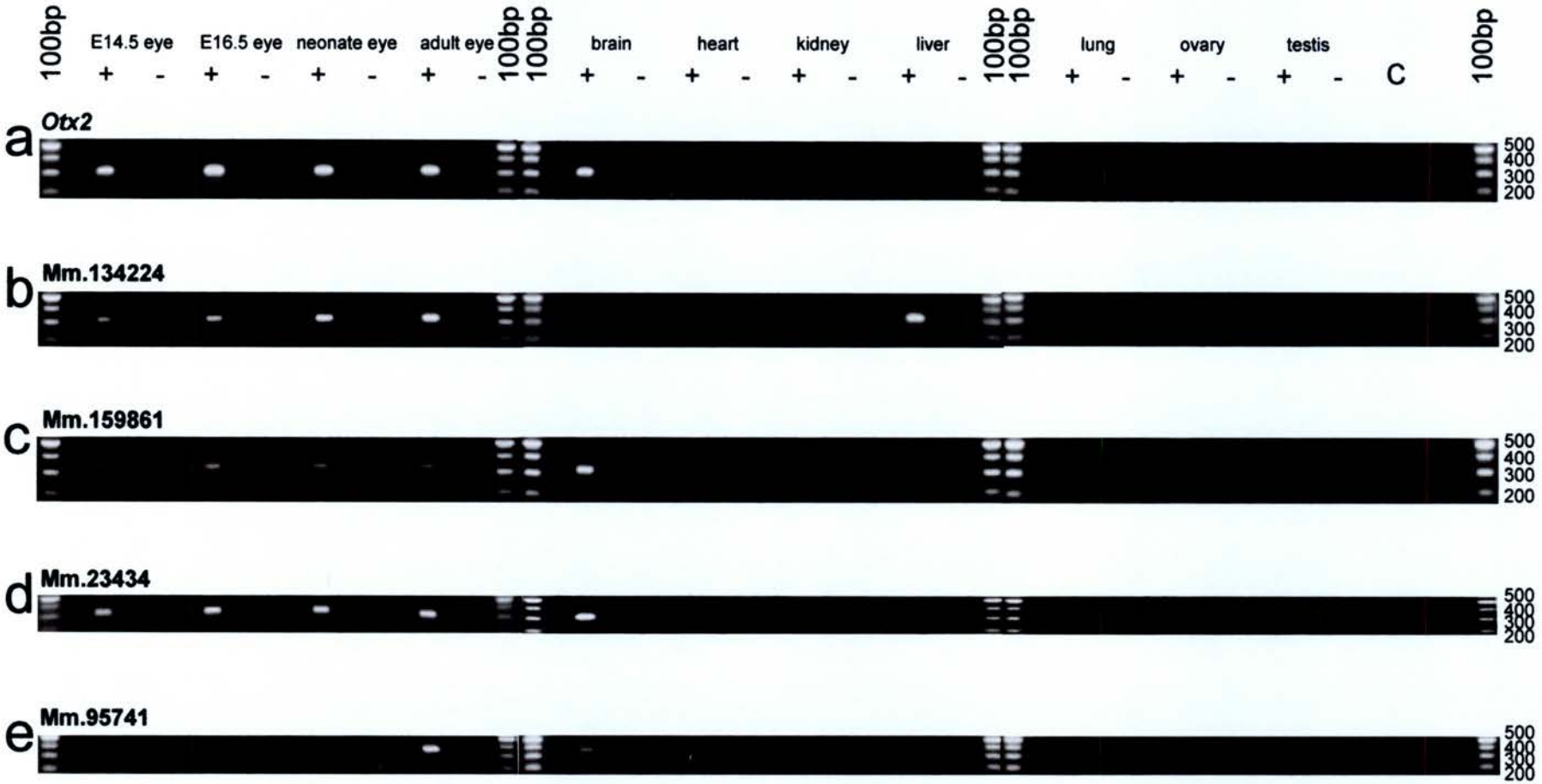


Figure 4.3 RT-PCR panels of clusters showing preferential expression in the eye
See next page for full legend.

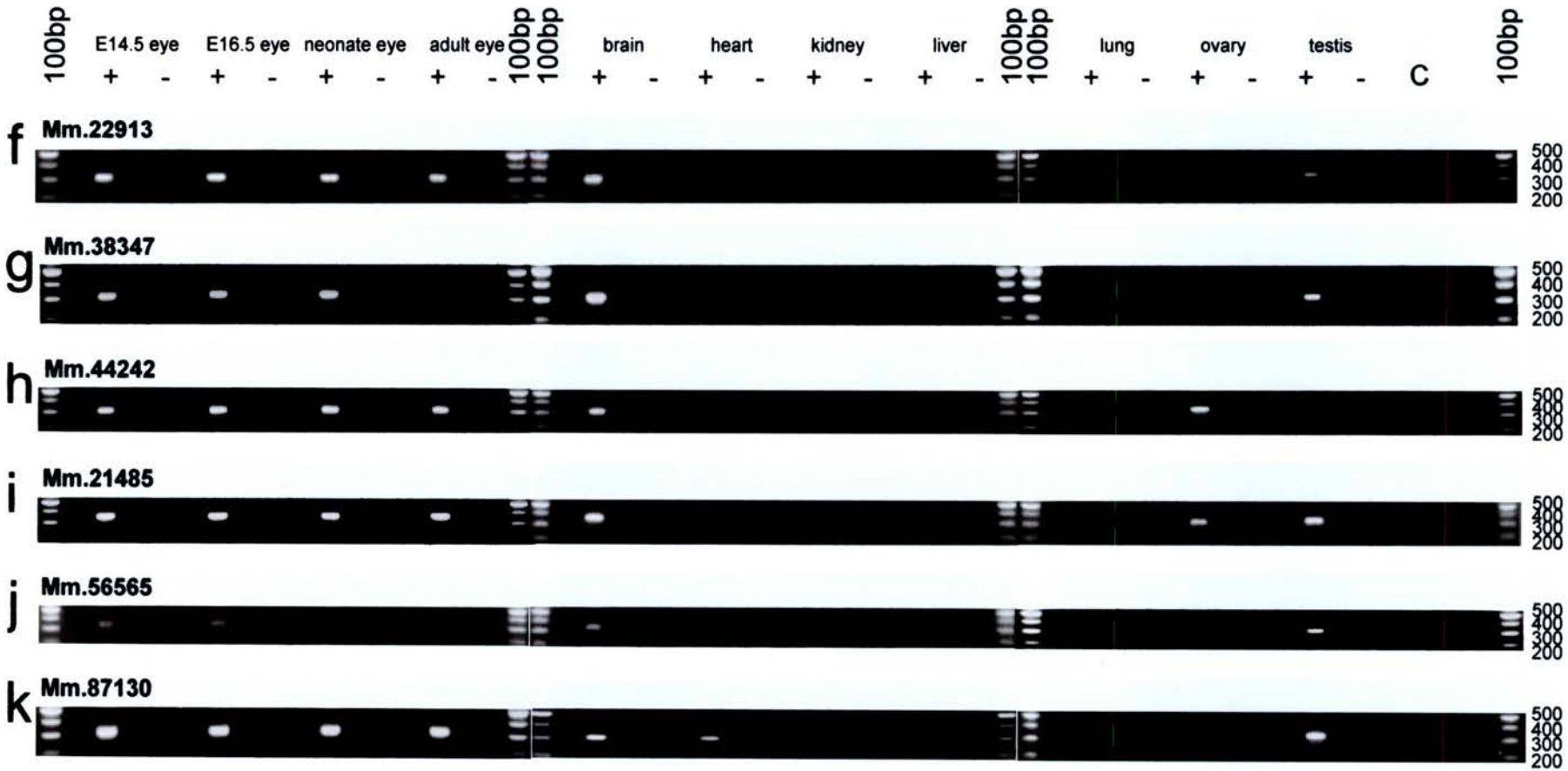


Figure 4.3 RT-PCR panels of clusters showing preferential expression in the eye
Eleven clusters show preferential eye expression, with five expressed in one non-eye tissue (a, b, c, d, e), three expressed in two non-eye tissues (f, g, h) and three expressed in three non-eye tissues (i, j, k). All except one (b) show expression in the brain and the same is true for testis expression (h is the exception). Testes and ovary expression are commonly seen in clusters expressed in more than one non-eye tissue. For each tissue, a reaction using a reverse transcription with no enzyme was carried out (-) in addition to the experimental reaction (+). An additional negative control with water is shown (C). All non-eye tissues are adult in origin. The number of base pairs represented by bands in the 100bp marker is shown on the right.

Expression classification	UniGene cluster	Represented gene	E14.5 eye	E16.5 eye	Neonate eye	Adult eye	Other expression
Eye specific	Mm.150838	-	●	●	●	○	-
	Mm.195923	*Crygf*	●	●	●	●	-
	Mm.59151	*Guca1b*	○	●	○	●	-
	Mm.95707	*Aipl1*	○	○	○	●	-
Preferentially expressed	Mm.134224	-	●	●	●	●	Li
	Mm.134516	*Otx2*	●	●	●	●	B
	Mm.159861	-	●	●	●	●	B
	Mm.23434	-	●	●	●	●	B
	Mm.95741	-	○	○	○	●	B
	Mm.22913	-	●	●	●	●	B,T
	Mm.38347	-	●	●	●	●	B,T
	Mm.44242	-	●	●	●	●	B,O
	Mm.21485	-	●	●	●	●	B,O,T
	Mm.56565	-	●	●	●	●	B,T
	Mm.87130	-	●	●	●	●	B,H,T
Non-preferentially expressed	Mm.103712	-	○	○	○	○	T
	Mm.121647	*Rdh12*	●	●	●	●	B,H,K,O,T
	Mm.200734	*D15Ertd806e*	●	●	●	●	B,H,Li,Lu,T
	Mm.213208	-	●	●	●	●	B,H,K,O,T
	Mm.29846	*Ndr4*	○	●	●	●	B,H,K,Lu,O,T
	Mm.3124	-	○	●	●	●	B,H,K,Lu,O
	Mm.137178	*Gabt4*	●	●	●	●	A
	Mm.151918	-	●	●	●	●	A
	Mm.2965	-	●	●	●	●	A
	Mm.34901	*Grtp1*	●	●	●	●	A
	Mm.35511	-	●	●	●	●	A
	Mm.37819	*D5Bwg0860e*	●	●	●	●	A
	Mm.82341	-	●	●	●	●	A
No detected expression	Mm.44292	-	○	○	○	○	-
	Mm.62589	-	○	○	○	○	-
	Mm.78118	-	○	○	○	○	-

Table 4.1 Summary of RT-PCR results
RT-PCR results are shown for all clusters tested and are categorised according to this outcome. Gene names are given for those clusters that were subsequently identified as representative of known genes. Filled circles represent a positive result with RT-PCR while unfilled circles represent a negative result. Expression in other tissues is also represented. B, brain; H, heart; K, kidney; Li, liver; Lu, lung; O, ovary; T, testis; A, all tissues.

The fifth and last known gene (Mm.134516, *Otx2*) is preferentially eye expressed and is one of five of these clusters that display expression in just one non-eye tissue.

For four of these five (*Otx2*, Mm.159861, Mm.23434 and Mm.95741), this tissue is the brain, with the other cluster (Mm.134224) being expressed in the liver. Interestingly, amongst the preferentially expressing clusters, ten are expressed in the brain, with Mm.134224 being the only exception. Ovary and testis also appear to express many of these clusters.

The three clusters not displaying expression in any of the tested tissues are Mm.44292, Mm.62589 and Mm.78118. Of the remaining 13 non-preferentially expressed genes, only Mm.103712 shows no eye expression, while seven are expressed in all tissues.

The results of this analysis have further refined the number of studied clusters, with 15 of 31 (48%) clusters shown to be specifically or preferentially expressed in the mouse eye and used as a basis for further investigation.

4.3 Analysis and preferential expression

In order to gain expression information at stages earlier than had been tested using RT-PCR, wholemount *in situ* hybridisation was performed on four embryonic stages (E9.5, E10.5, E11.5 and E12.5) using the 12 novel clusters chosen for further analysis. *In situ* hybridisation was also performed on cryosections at later stages (E14.5 and E16.5 CD1 embryos and neonate and adult CD1 eyes) in an effort to elucidate cell locations and types that express the novel clusters discussed in this thesis, as well as to confirm RT-PCR expression of these clusters. A DOPAchrome tautomerase (*Dct*) specific probe was employed as a positive control for eye expression. Two of the known clusters (Mm.134516, *Otx2*; Mm.195923, *Crygf*) also had their early embryonic expression assessed as an indication of the desired activity of probes designed and synthesised using the methods described in this thesis. All

sense and antisense probes were synthesised from the same plasmid, in which inserts were cloned in either direction. The restriction enzymes used to linearise the plasmids were SpeI and NcoI, with the exception of the plasmid containing the Mm.22913 insert, when SacII was used due to the presence of an NcoI site in the insert. *In situ* analysis carried out on the novel clusters detected expression for nine of the 12 tested.

In order to elucidate possible roles for these eye-expressing genes, a bioinformatic approach was taken. This was initiated by BLAST searches against the mouse genome using the Ensembl interface (www.ensembl.org). Predicted protein information was used as a basis for further investigation in some cases.

The mouse regions and the homologous human regions for the nine novel clusters described in this chapter are shown (Table 4.2). There are no existing gene traps of any of the cluster sequences in the available gene trap databases.

UniGene cluster	Proposed name	Mouse region	Human region	Linked disease
Mm.150838	-	17E4	2p21	-
Mm.159861	*Tsd*	1B	6p12.1	autosomal dominant congenital
Mm.21485	*Tset*	10B1	6q22.1	-
Mm.22913	*Trale*	15E3	22q13.31	-
Mm.23434	*Cln8l*	7F4	16p11.2	-
Mm.38347	*Bret*	11E2	17q25.1	age-related maculopathy
Mm.44242	*Sapan*	5B1	4p16.1	-
Mm.87130	*Duf676*	1A5	6q13	macular dystrophies
Mm.95741	-	2A2	10p12.31	-

Table 4.2 Regions and mapped diseases of novel clusters
The UniGene cluster, proposed gene name, mouse region and homologous human region are shown along with relevant diseases mapped to these regions. *Tsd*, antisense to *Dst*; *Tset*, *Tspy/Set* related; *Trale*, transmembrane, retina and antisense lens expression; *Cln8l*, *Cln8*-like; *Bret*, BTB/POZ domain, retinal; *Sapan*, saposin A domain, novel; *Duf676*, DUF676 domain.

The twelve clusters discussed here describe a sequence with no clear orthologues (Mm.95741), a sequence upstream of *Six3* with retinal expression (Mm.150838), a

sequence that lies antisense to an exon in a known gene (*Tsd*), a novel gene displaying antisense expression (*Trale*) and five novel genes with interesting expression patterns and predicted protein domains with known functions (*Tset*, *Bret*, *Cln8l*, *Sapan* and *Duf676*).

4.3.1 Control probes: *Dct*, *Otx2* and *Crygf*

The *Dct* antisense riboprobe was synthesised from the p5A7 plasmid and the sense probe was transcribed from the p5A3 plasmid. These probes have been previously used to detect *Dct* expression in the developing eye (Steel et al., 1992). Expression is shown in the RPE at E12.5 as expected (Figure 4.5) with no expression in the sense control. Expression can also be seen in melanoblasts and the telencephalon. CD1 (albino) embryos were used in all *in situ* hybridisation experiments so that pigmentation did not mask a retinal signal. Expression of *Dct* is also shown at E14.5, E16.5 and neonate stages (Figure 4.6). Signal is seen in the RPE at all stages and in the IPE and ciliary epithelium in the neonate as expected. Expression can also be seen in melanoblasts at E14.5. Retinal detachment has taken place in the E16.5 section, and in many of the later sections, as a consequence of tissue processing.

The *Otx2* probe shows expression in the retina of the developing eye at E11.5 (Figure 4.7) and E12.5 (Figure 4.8). Expression is also seen in the brain at both stages. This was expected, as vertebrate expression of *Otx2* has previously been reported in the telencephalon, diencephalon and mesencephalon (Simeone et al., 1992) and eye expression at E11.5 and E12.5 is known to be restricted to the RPE (Bovolenta et al., 1997).

As *Crygf* is expressed in the lens, it presents the opportunity to assay expression in a non-retinal region of the eye. This is shown at E10.5, E11.5 and E12.5 (Figure 4.9).

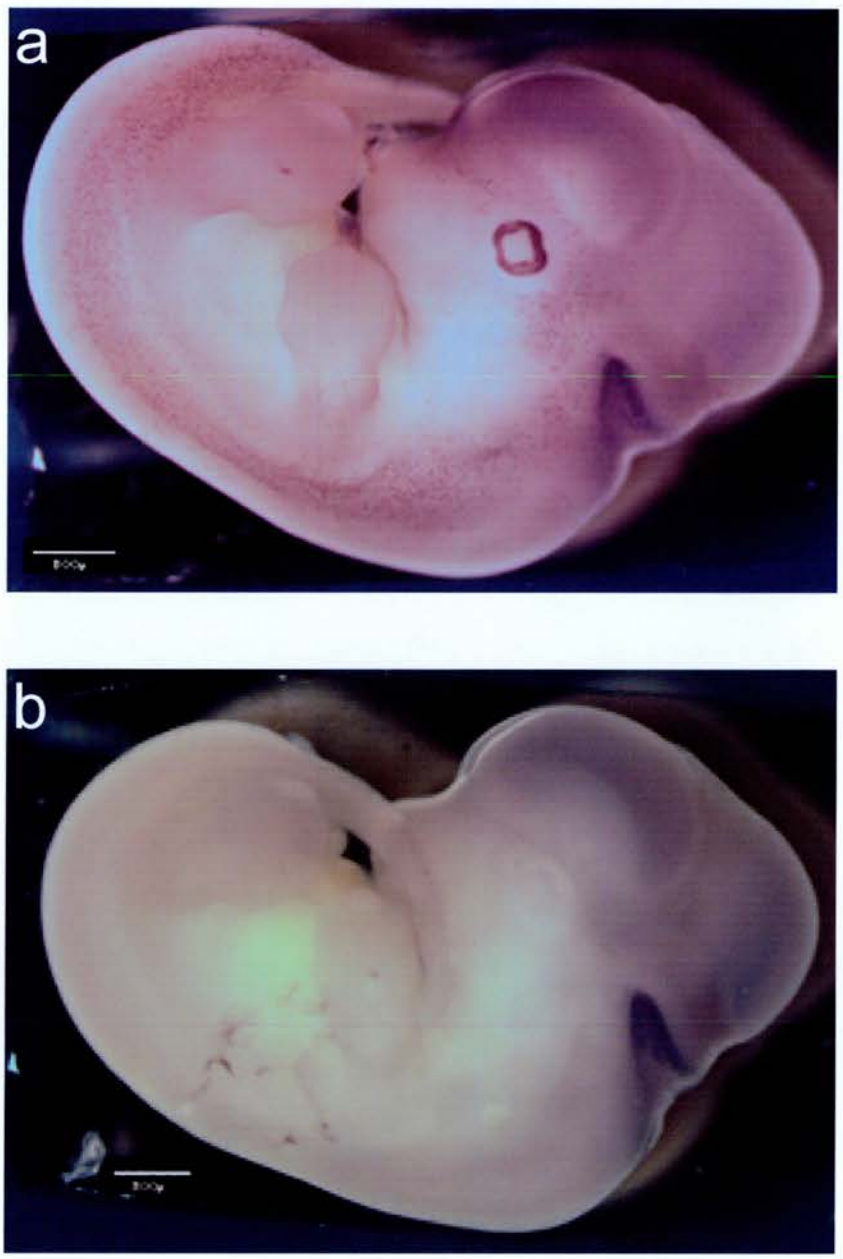


Figure 4.5 *Dct* wholemount *in situ* hybridisation
In situ hybridisation at E12.5 using *Dct* antisense probe shows expression in the RPE and melanoblasts (a). The sense control is also shown (b). Scale bars, 800 μm.

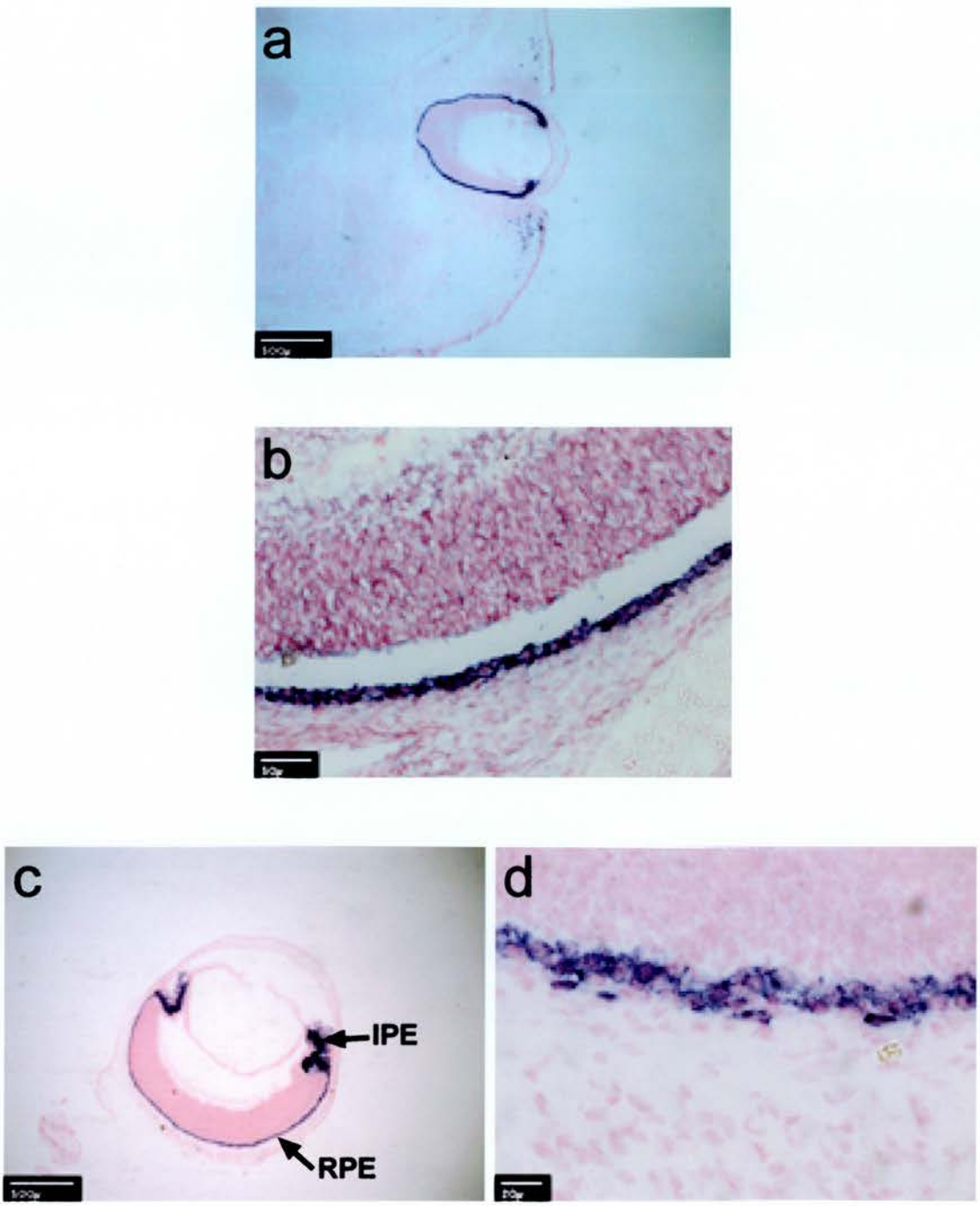


Figure 4.6 *Dct in situ* hybridisation
In situ hybridisation on sagittal sections. At E14.5, expression can be seen in melanoblasts and in the RPE and IPE of the eye (a). At E16.5, expression is visible in the RPE, although it is detached from the neural retina due to sectioning procedures (b). The cilary body, IPE and RPE are all expressing in the neonate (P1) eye (c) and the RPE is shown at higher magnification (d). High magnification images show the choroid/lens axis from the bottom to the top of the image. RPE, retinal pigment epithelium; IPE, iris pigment epithelium. Scale bars, 20 μm (d); 50 μm (b); 500 μm (a,d).

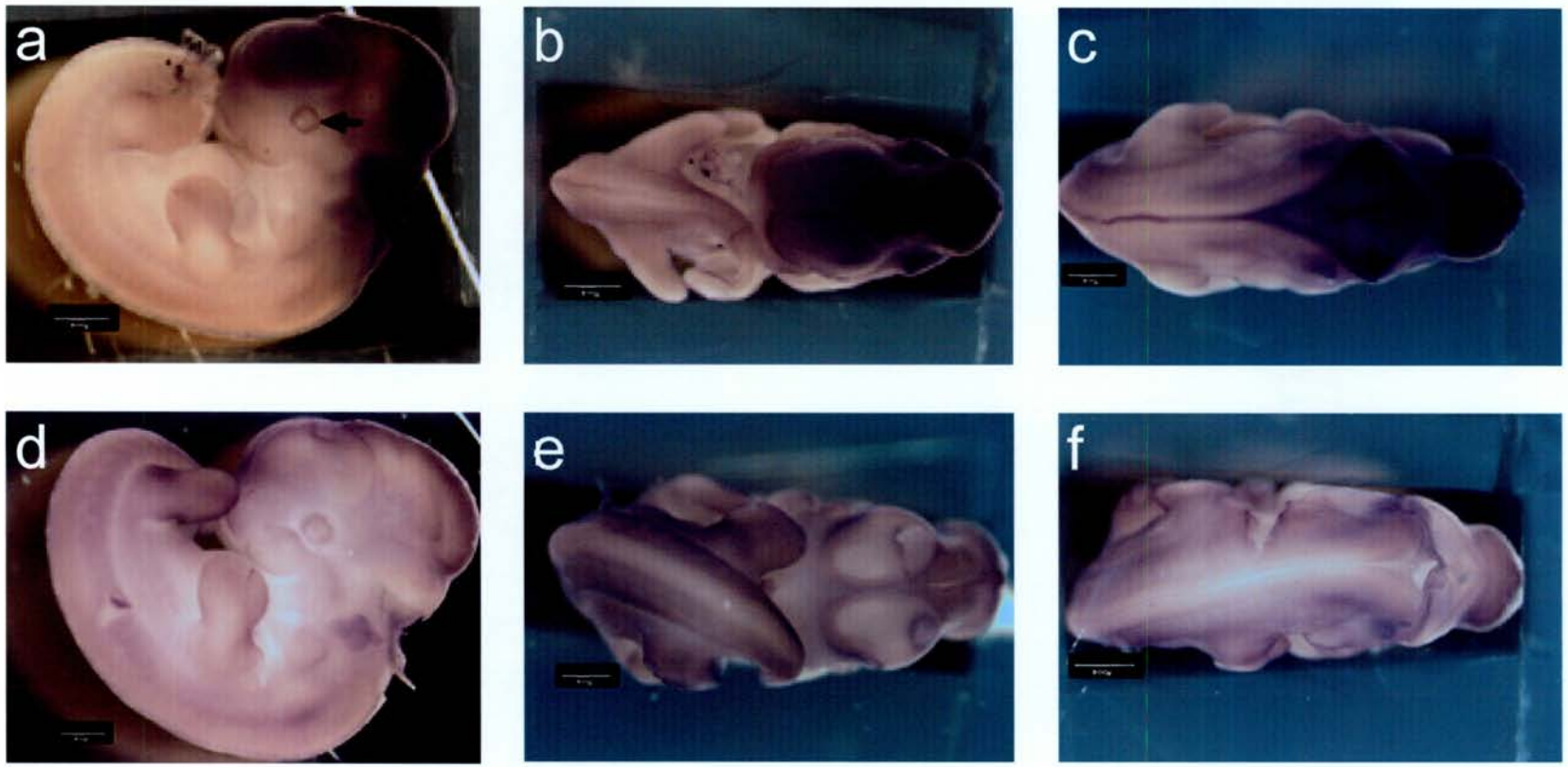


Figure 4.7 *Otx2* wholemount *in situ* hybridisation at E11.5
The lateral view of *Otx2 in situ* hybridisation shows expression in the retina (a, experimental, indicated by arrow; d, control). Additional signal is visible in the developing brain, also shown in ventral (b, experimental; e, control) and dorsal (c, experimental; f, control) views. Scale bars, 800 μm.

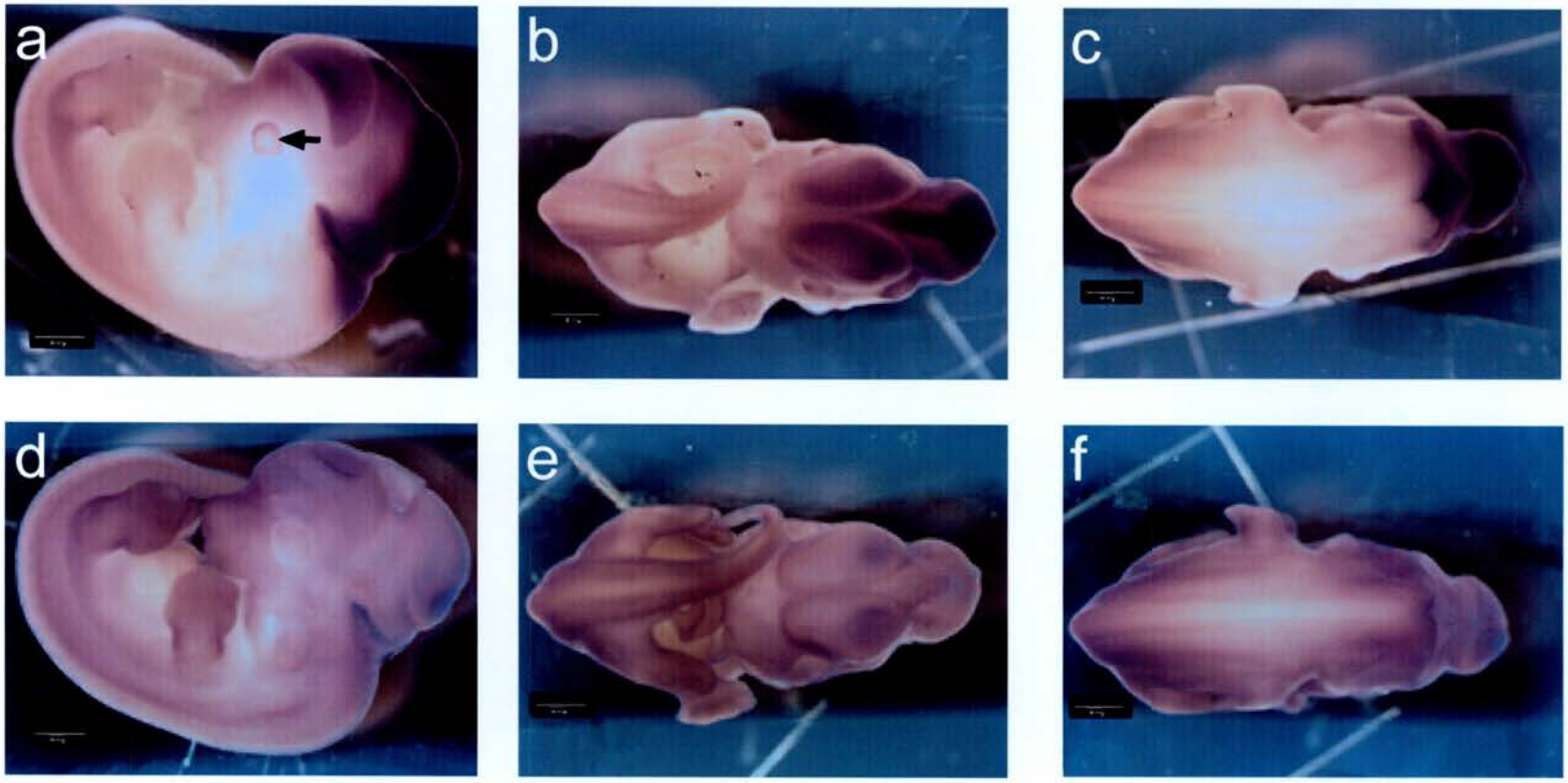


Figure 4.8 *Otx2* wholemount *in situ* hybridisation at E12.5
The lateral view of *Otx2 in situ* hybridisation shows expression in the retina (a, experimental, indicated by arrow; d, control). Additional signal is visible in the developing brain, also shown in ventral (b, experimental; e, control) and dorsal (c, experimental; f, control) views. Scale bars, 800 μm.

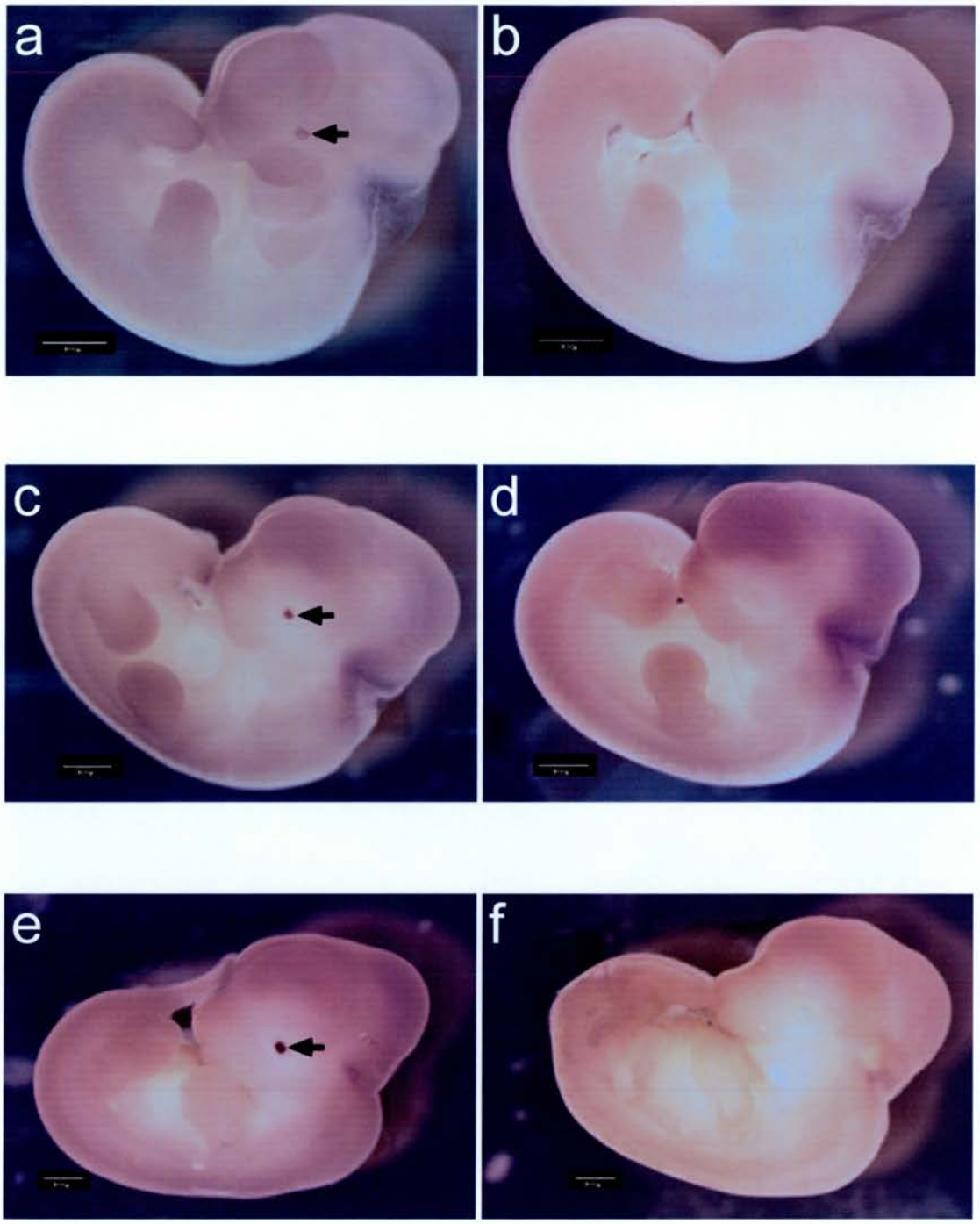


Figure 4.9 *Crygf* wholemount *in situ* hybridisation
In situ hybridisations at E10.5 (a, experimental; b, control), E11.5 (c, experimental; d, control) and E12.5 (e, experimental; d, control) show expression in the developing lens (indicated by arrows). Scale bars, 800 μm.

4.3.2 Mm.95741

RT-PCR results suggest that Mm.95741 is only expressed in the adult eye and not in the late embryo at E14.5, E16.5 or in the neonate. Wholemount *in situ* hybridisations reveal that Mm.95741 is expressed in the early embryo. At E9.5 expression is restricted to the optic and otic vesicles and the brain and at E10.5 less pronounced signal is seen in the retina, otic vesicle and brain (Figure 4.10). However, *in situ* hybridisation at stages E11.5 and E12.5 showed no expression. This was also the case with *in situ* analysis at E14.5, E16.5 and in the neonate eye as well as in the adult eye, which is contrary to the earlier RT-PCR data (section 4.2).

The representative sequence of Mm.95741 is AK020807, which aligns to the last exon of a Genscan predicted mouse gene (Figure 4.11). This prediction links a single exon gene (2810030E01Rik) to the Mm.95741 sequence, as they align with consecutive predicted Genscan exons. The 2810030E01Rik gene has homologues in human (Q8N3C3), rat (novel), zebrafish (novel) and fugu (Q9CSH2), but has no distinguishing domains or features. Mm.95741 aligns with a region downstream of the human and rat genes as well as the mouse prediction, however it is only in zebrafish that there is an EST in the corresponding region.

Although there are no ESTs seen in the human or rat regions, the sequence representing Mm.95741 lies downstream of a gene that has orthologues in other vertebrates. Mm.95741 may represent a conserved region syntenic to this gene or it may be its 3' UTR. However, there is no evidence to suggest that this gene is preferentially expressed in the eye, thus Mm.95741 could be an alternative 3' UTR, preferential to the developing eye, ear and brain, and possibly adult eye and brain according to RT-PCR data.

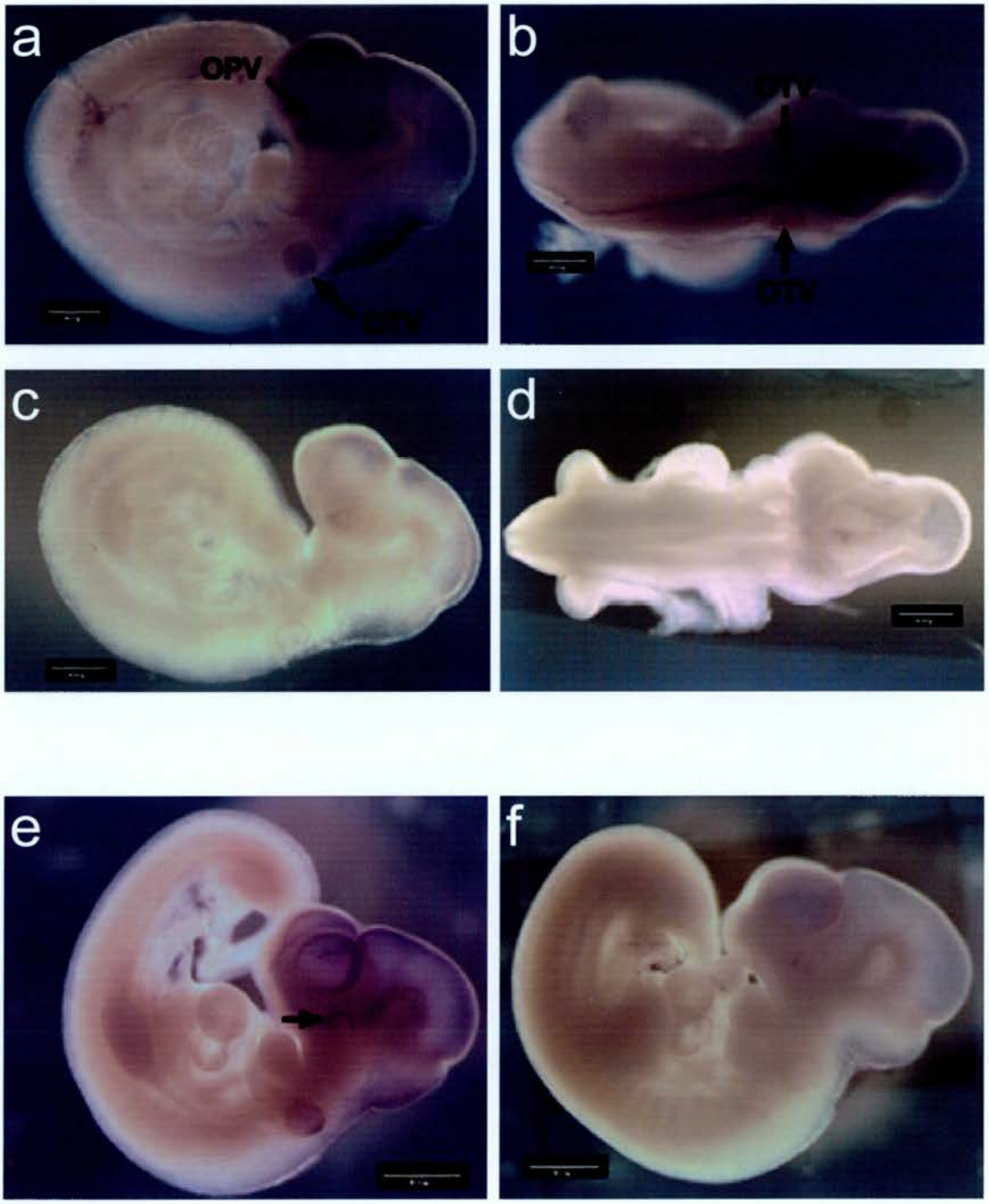


Figure 4.10 Mm.95741 wholemount *in situ* hybridisation
The lateral view of Mm.95741 *in situ* hybridisation at E9.5 shows expression in the optic and otic vesicles and the brain (a, experimental; c, control). Brain and otic vesicle expression can also be seen in the dorsal view (b, experimental; d, control). At E10.5, expression remains visible in the brain and to a lesser extent in the otic vesicle (e, experimental; f, control). Within the developing eye, signal is restricted to the retina (indicated by arrow). OPV, optic vesicle; OTV, otic vesicle. Scale bars, 400 μm (a, b, c, d); 800 μm (e, f).

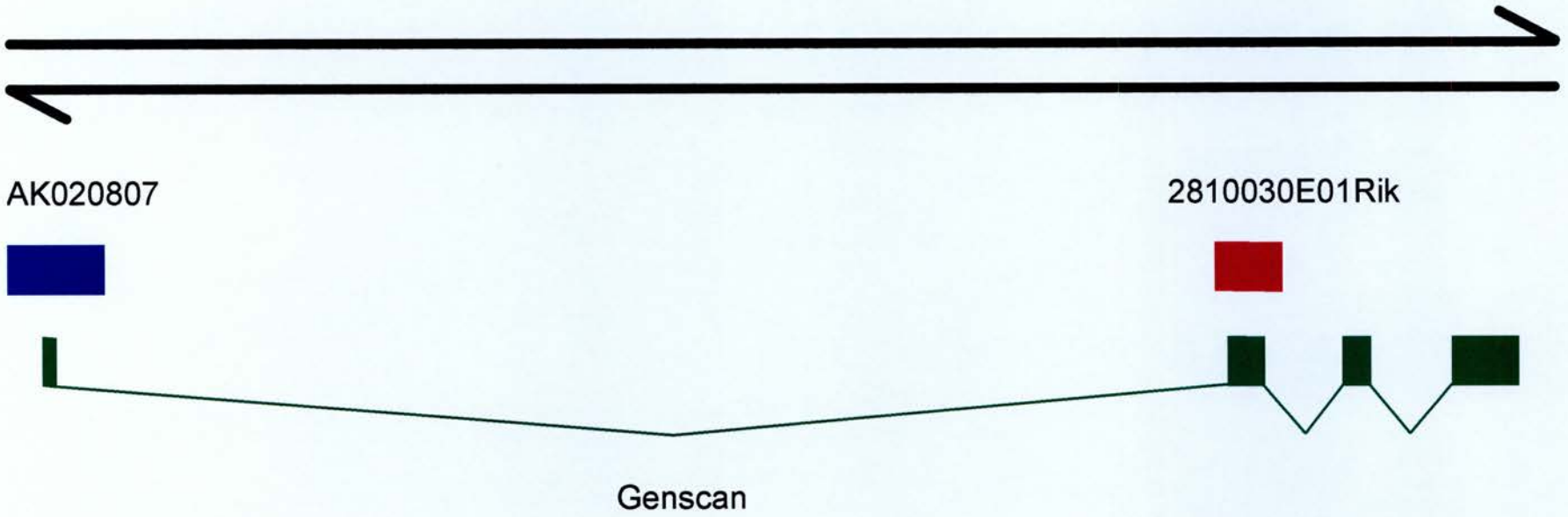


Figure 4.11 The mouse Mm.95741 region
AK020807, the representative sequence of Mm.95741, aligns to an exon of a predicted Genscan transcript. This prediction also has an exon corresponding to a gene (2810030E01Rik) with homologues in human, rat, zebrafish and fugu. Not to scale.

4.3.3 Mm.150838

RT-PCR results for Mm.150838 indicate specific eye expression in the tissues used. Expression is only seen in the eye while it is undergoing development (E14.5, E16.5 and neonate) but not in the adult. Expression of Mm.150838 in the neonate eye is confirmed by section *in situ* hybridisation (Figure 4.12). The retinal cells display signal, with the inner nuclear layer and discrete cells, lying in a particular retinal layer on the peripheral side of the central retinal region, showing the strongest expression. High expression is also detected in a layer of cells that may correspond to the innermost choroidal layer or the RPE and also in the RGCs. Although *in situ* hybridisation was carried out at all stages, no expression was seen at stages E9.5, E10.5, E11.5, E12.5 or in the adult. Expression was also absent from the eye at stages E14.5 and E16.5, contradicting the RT-PCR results.

In both mouse and human, AK021382, the representative sequence for Mm.150838, aligns with the exon of an EST sequence within a group of ESTs that lie 7kb upstream of, and in the opposite sense to, *Six3*. In order to assess the homology between human and mouse in this region, VISTA analysis was performed (Figure 4.13). The exonic similarity to Mm.150838 is from the second and final exon of ENSESTT00000601866, which shows between 50% and 75% homology between the two organisms, whilst the other two exons that are part of the same cDNA sequence (in ENSESTT00000601864) display over 75% homology. In addition, there are regions of over 75% homology between the 5' regions of *Six3* and these ESTs. However, Mm.150838 has no ORF.

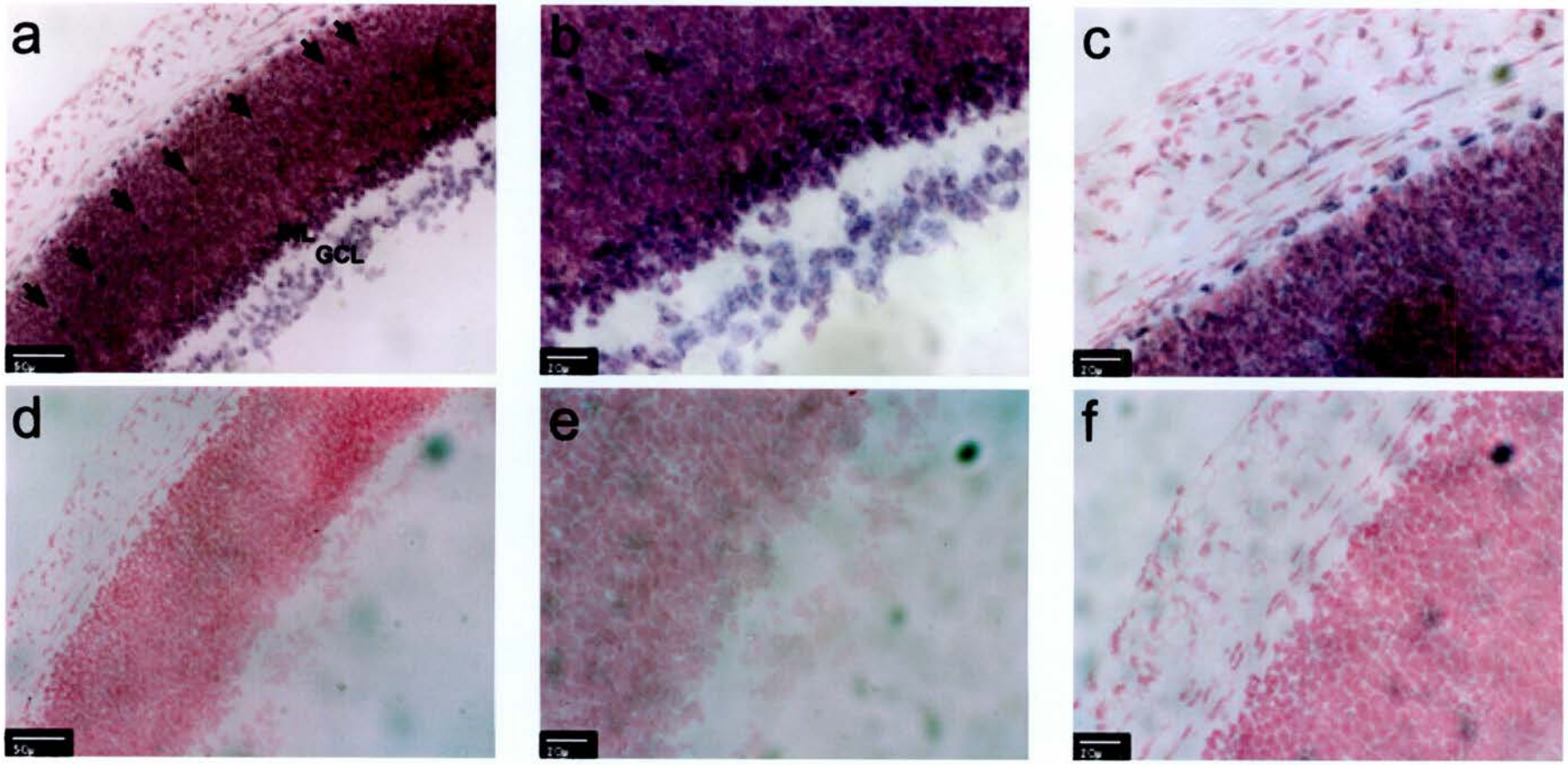


Figure 4.12 Mm.150838 *in situ* hybridisation
In situ hybridisation on sagittal sections of neonate (P1) retina show signal throughout the retina and in the layer of the choroid that lies adjacent to the RPE (a, experimental; d, control). Cells showing strong signal are distributed in a layer that lies on the peripheral side of the central retina (indicated by arrows). These cells are visible in the higher magnification images. Inner retinal expression is shown more clearly in a higher magnification image, including the strongly expressing INL cells (b, experimental; e, control). A higher magnification image of the outer retinal region shows choroidal or RPE expression more clearly (c, experimental; f, control). Images show the choroid/lens axis from top left to bottom right. INL, inner nuclear layer; GCL, ganglion cell layer. Scale bars, 20 μm (b, c, e, f); 50 μm (a, d).

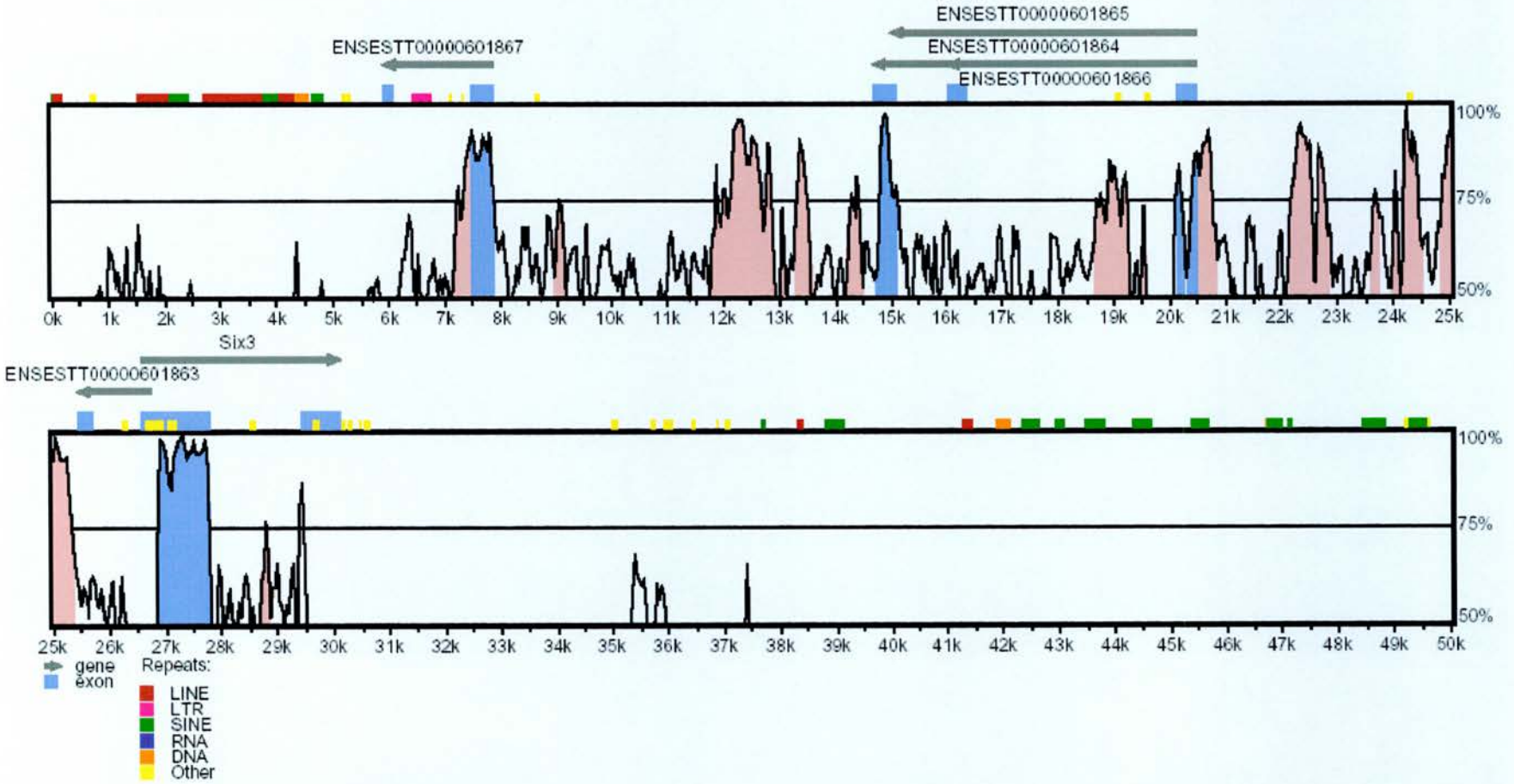


Figure 4.13 VISTA plot of the Mm.150838 region
Shaded areas represent regions of over 75% homology between human (base sequence) and mouse, both exonic (blue) and regions with no known exons (pink). The representative sequence of Mm.150838 aligns with the second exon of ENSESTT00000601866, with homology in this region being less than 75%. There are regions of over 75% homology between the 5' regions of *Six3* and the cluster of 3 ESTs.

The expression of Mm.150838 in neonate retina is widespread, with signal throughout the retina and also in a single layer that may correspond to a layer of the choroid or RPE. The retinal cells with more intense Mm.150838 staining are those in the INL (bipolar cells) and the line of dispersed cells that show a similar expression pattern to that of *Prox1* in differentiating horizontal cells at the same stage (Dyer et al., 2003). Therefore, Mm.150838 may be important in retinal development or function, either indirectly or directly. As the representative sequence of Mm.150838 has no ORF, it may represent an untranslated part of a transcript in this region, as suggested by the existence of ESTs located here. The presence of conserved regions between the Mm.150838 gene and *SIX3* suggests a regulatory role for these regions on the two genes' expression. The non-exonic conserved region immediately upstream of *SIX3* (surrounding 25kb in Figure 4.13) has been previously identified as a possible promoter region (Lengler and Graw, 2001), whilst other regions have not been reported in the literature. Clustering of tissue specific genes has been reported in the literature as occurring more often than would be expected by chance (Megy et al., 2003). If this is a genuine phenomenon, with implications for shared regulation of gene expression at some level, Mm.150838 and *Six3* may be subject to the same influences. This hypothesised relationship is supported by the role and expression of *Six3*. Mutations in the human orthologues cause holoprosencephaly type 2 (Wallis et al., 1999) as it is a homeodomain protein required for eye formation (Lagutin et al., 2001). In the neonate mouse retina (Zhu et al., 2002) and human retina at 20 weeks gestation (Granadino et al., 1999), *SIX3* is expressed in RGCs and the INL in the neonate, a subset of the expression seen with Mm.150838.

4.3.4 *Tsd*

RT-PCR detects expression of *Tsd* in the eye at all tested stages (E14.5, E16.5, neonate and adult) as well as in the brain. *In situ* hybridisation at earlier stages of development detected expression of *Tsd* at E9.5 and E10.5 (Figure 4.14). At E9.5, there appears to be widespread expression, including expression in the optic cup. However, at E10.5, both the antisense and sense probe appear to show expression in the retina of the optic cup. The experimental probe also detects possible expression in the brain, whilst the control probe detects expression exclusively in the optic cup. Although eye expression was seen at all stages tested by RT-PCR (E14.5, E16.5, neonate and adult), the *in situ* results at these stages showed no expression. Expression was also absent from *in situ* analysis at E11.5 and E12.5. The brain expression indicated by adult brain RT-PCR has not been shown to be present at embryonic stages.

In both human and mouse, AK020813, the representative sequence of Mm.159861, is located antisense to dystonin (*Dst*; bullous pemphigoid antigen 1 (*BPAG1*) in human). In the human genome, it comprises two exons, the first of which overlaps the second of two exons of an EST (ENSESTT00002995780) (Figure 4.15). Therefore, taken together, these sequences probably represent a transcript deriving from three exons. If this is the case, the third exon is the only one not to have a corresponding antisense *BPAG1* exon. A vista plot of the relevant region shows that whilst the first two exons show homology between human and mouse, the third exon, the second of AK020813, does not (Figure 4.16). Additionally, the protein sequence predicted for the human EST has no domains or features and the representative sequence of Mm.159861 has no significant ORF. In zebrafish, the sequence aligns antisense to an exon of an unannotated gene.

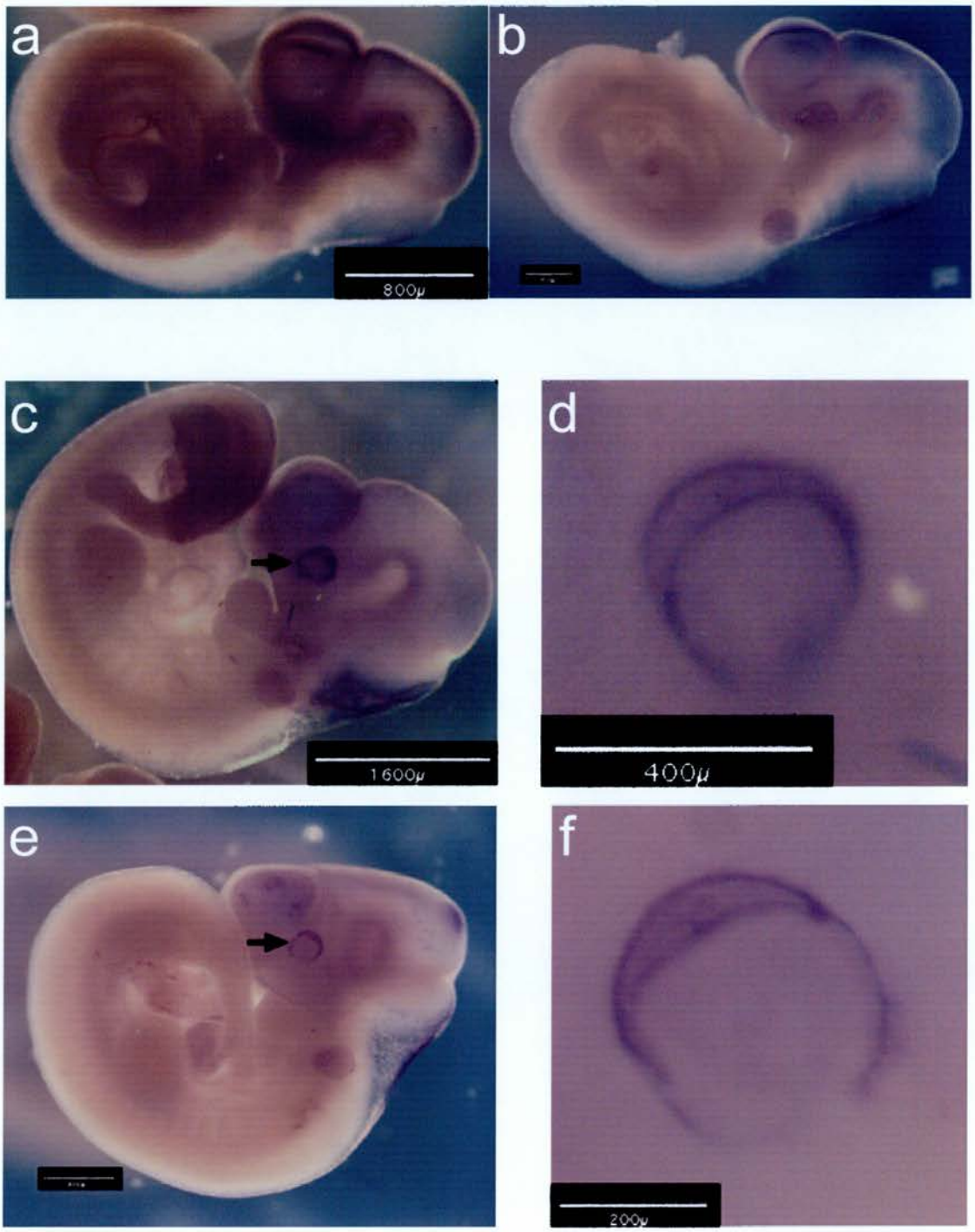


Figure 4.14 *Tsd* wholemount *in situ* hybridisation
At E9.5, expression is ubiquitous, including the optic vesicle, with the experimental probe (a), while the control probe shows relatively little expression (b). At E10.5, both the experimental (c) and the control (e) probes show expression in the developing retina (indicated by arrows), with the sense also showing brain signal. Higher magnification images display the eye nasal to temporal from bottom left to top right of the image (d, experimental; f, control) . Scale bars, 200 μm (f); 400 μm (b, d); 800 μm (a, e); 1.6 mm (c).

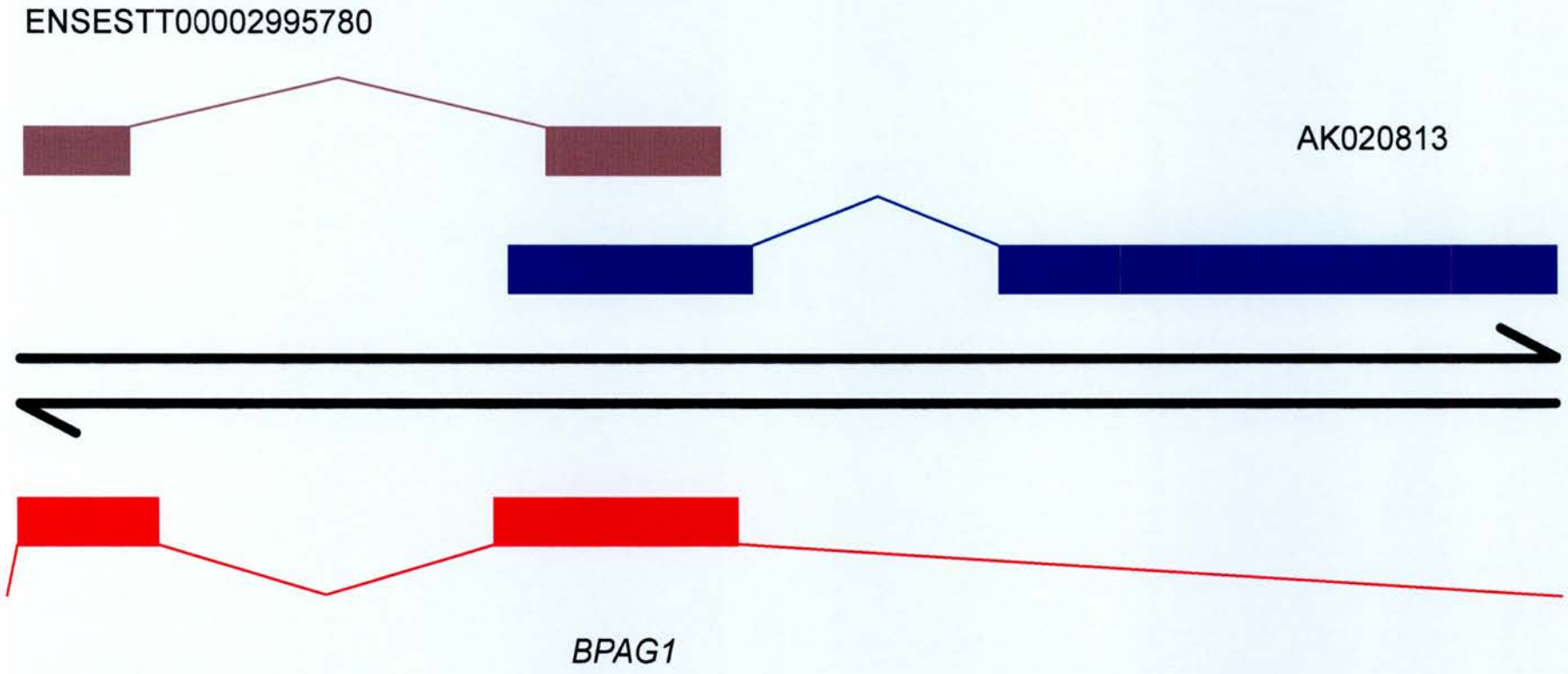


Figure 4.15 The human *Tsd* region
AK020813, the representative sequence of Mm.159861, is located antisense to *BPAG1*, with one antisense exon. An EST (ENSESTT00002995780) is present in the same region, overlapping one exon. Not to scale.

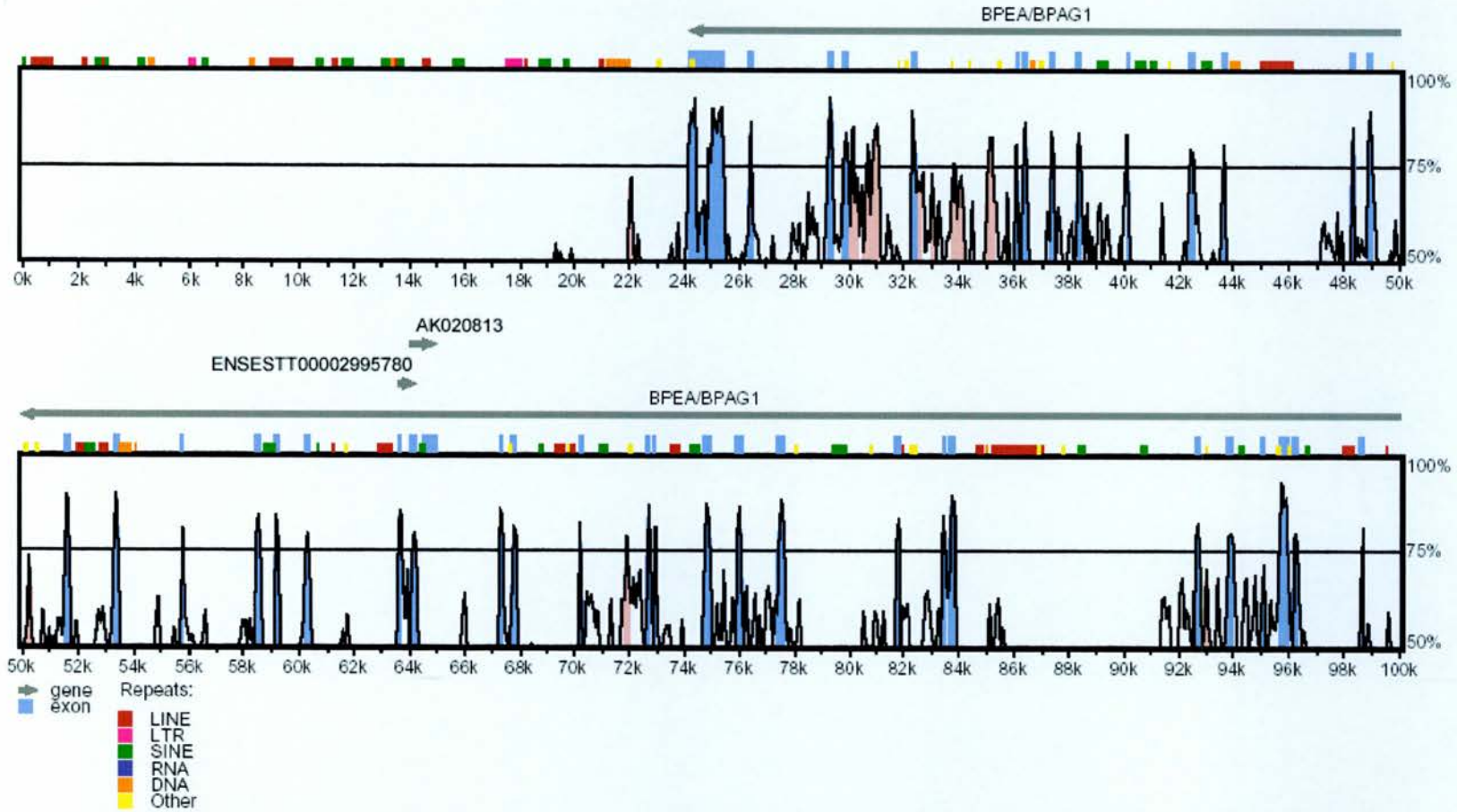


Figure 4.16 VISTA plot of the *Tsd* region

Shaded areas represent regions of over 75% homology between human (base sequence) and mouse, both exonic (blue) and regions with no known exons (pink). The two exons of ENSESTT00002995780 are antisense to two exons of *BPAG1*. The Mm.159861representative sequence (AK020813) encompasses the second EST exon and a novel exon which shows no homology between mouse and human and is not antisense to a *BPAG1* exon.

BPAG1 has many isoforms, all producing spectroplakin proteins, which are cytoskeletal crosslinkers. A mutation in *BPAG1* is responsible for causing bullous pemphigoid, a human autoimmune disease (Minoshima et al., 1991) and spontaneous mutations in BPAG1 are associated with dystonia musculorum (Duchen et al., 1964; Brown et al., 1995), a disease characterised by degeneration of sensory neurons (Bernier et al., 1995; Dowling et al., 1997). The *in situ* hybridisation riboprobes for *Tsd* are complementary to the second exon of the representative sequence and do not correspond to a known *BPAG1* exon on the opposite strand. It is possible that *Tsd* may be antisense to an undiscovered exon. However, the lack of conservation specifically in this region between mouse and human makes this less likely. As the representative sequence of *Tsd* is homologous to a region in zebrafish that is antisense to the exon of an unannotated gene, it is possible that the function of *Tsd* has been conserved and that the novel zebrafish gene is the *Dst* homologue. The predicted exonic region corresponding to Mm.159861 may be the 3' UTR of the EST gene, which would explain the lack of conservation in this region specifically. This gene may have a regulatory function and so influence the transcription, elongation, processing, splicing, localisation, stability, translation, transport or mRNAi mediated degradation of *Dst* transcripts. Interestingly, a different antisense transcript to *Dst* has recently been discovered further upstream (Kiyosawa et al., 2003). The phenomenon of antisense expression is discussed further elsewhere (section 4.4).

An autosomal dominant form of the disorder congenital nystagmus maps to the homologous human region on 6p12 in a region spanning 28Mb and including over 90 known genes (Kerrison et al., 1996). Congenital nystagmus has onset at infancy with a reduced visual acuity due to slippage of images across the fovea as a result of ocular oscillations. However, the physiological basis of this has not yet been determined and there has been no evidence of an underlying visual sensory disorder, (Kerrison et al., 1996), therefore there is no evidence to suggest that BPAG1 or *Tsd* are likely candidates.

4.3.5 *Trale*

RT-PCR data for *Trale* (section 4.2) indicates expression at all tested eye stages (E14.5, E16.5, neonate and adult) and also in brain and testis. In wholemount *in situ* hybridisations, expression is seen in both sense and antisense directions, as was the case with *Tsd*. At E9.5 and E10.5, strong expression seems to be restricted to the otic vesicle in the antisense direction (Figure 4.17). However, at E11.5, expression is detected in the eye in both directions (Figure 4.18). In the sense direction, expression appears to be in the retina. In the antisense direction, there is no retinal expression but instead it is present in the lens vesicle. Section *in situ* hybridisations display signal in the neonate retina, but not at E14.5, E16.5 or in the adult as suggested by the RT-PCR results. Additionally, there is no expression at E12.5. In the neonate eye, signal can be seen with the experimental probe, but not with the control probe, in discrete points distributed throughout the retina (Figure 4.19). This may represent a background effect and not real expression, however if this is the case, it is specific to certain riboprobes, since the sense controls show no signal. This is possible, as signal at the periphery of sections, a phenomenon known as edge effect, also seems to be specific to particular riboprobes (not shown), however edge effect has been observed with both sense and antisense riboprobes whilst the non-specific signal has only occurred with antisense riboprobes.

In Ensembl, the representative sequence of Mm.22913 (AK007745) aligns to a Genscan mouse gene prediction that includes a novel mouse gene (ENSMUSG00000036118). The protein deriving from this novel gene is predicted to have a transmembrane domain but no other recognisable domains. There are homologues of this gene in human (Q9BYB5), rat (novel ENSRNOG00000011788), zebrafish (novel ENSDARG00000005569) and fugu (Q9BYB5). With the exception of the fugu homologue, which is only predicted to have a signal peptide, all of these genes have a predicted transmembrane domain. In Ensembl, none of these transcripts have antisense genes.

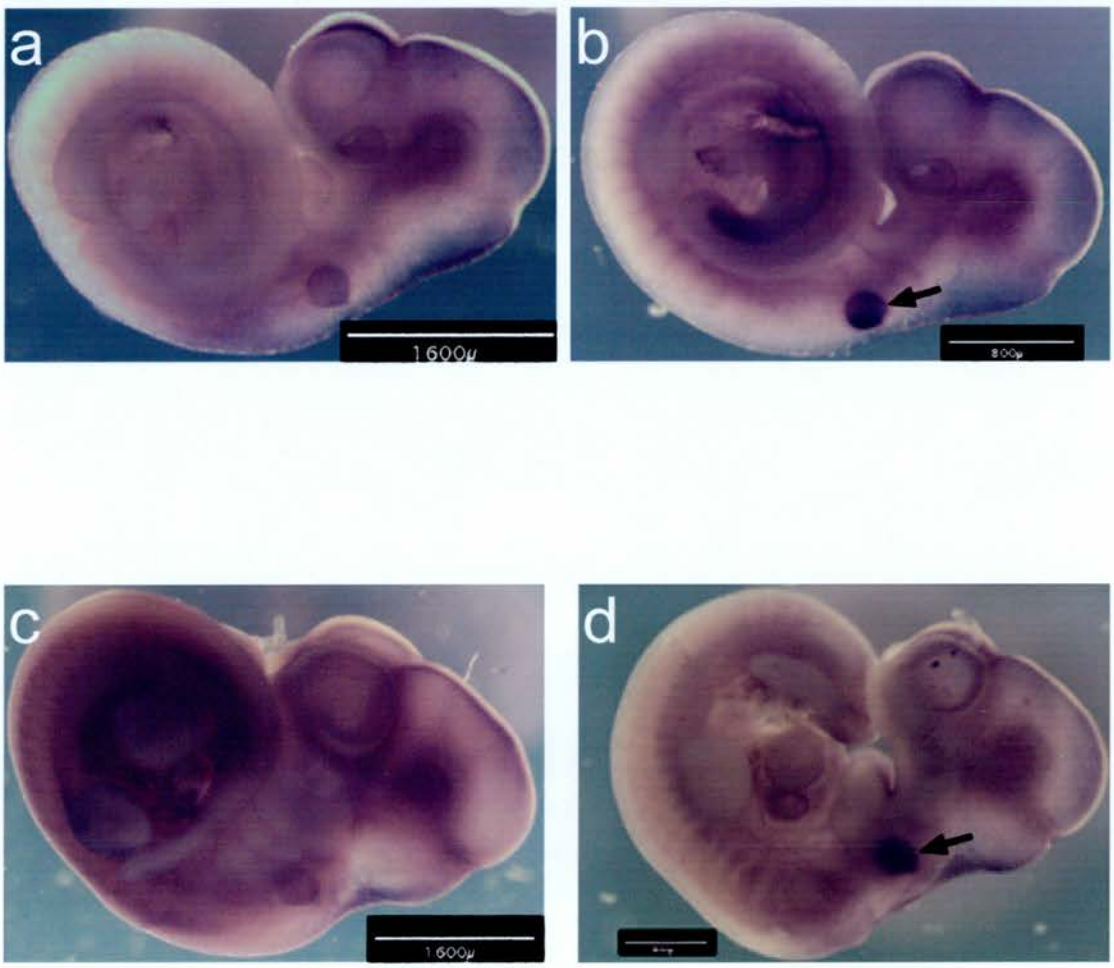


Figure 4.17 *Trale* wholemount *in situ* hybridisation at E9.5 and E10.5
At both E9.5 (a, experimental; b, control) and E10.5 (c, experimental; d, control) strong antisense expression is seen in the otic vesicle (indicated by arrows). Scale bars, 800 μm (b, d); 1.6 mm (a, c).

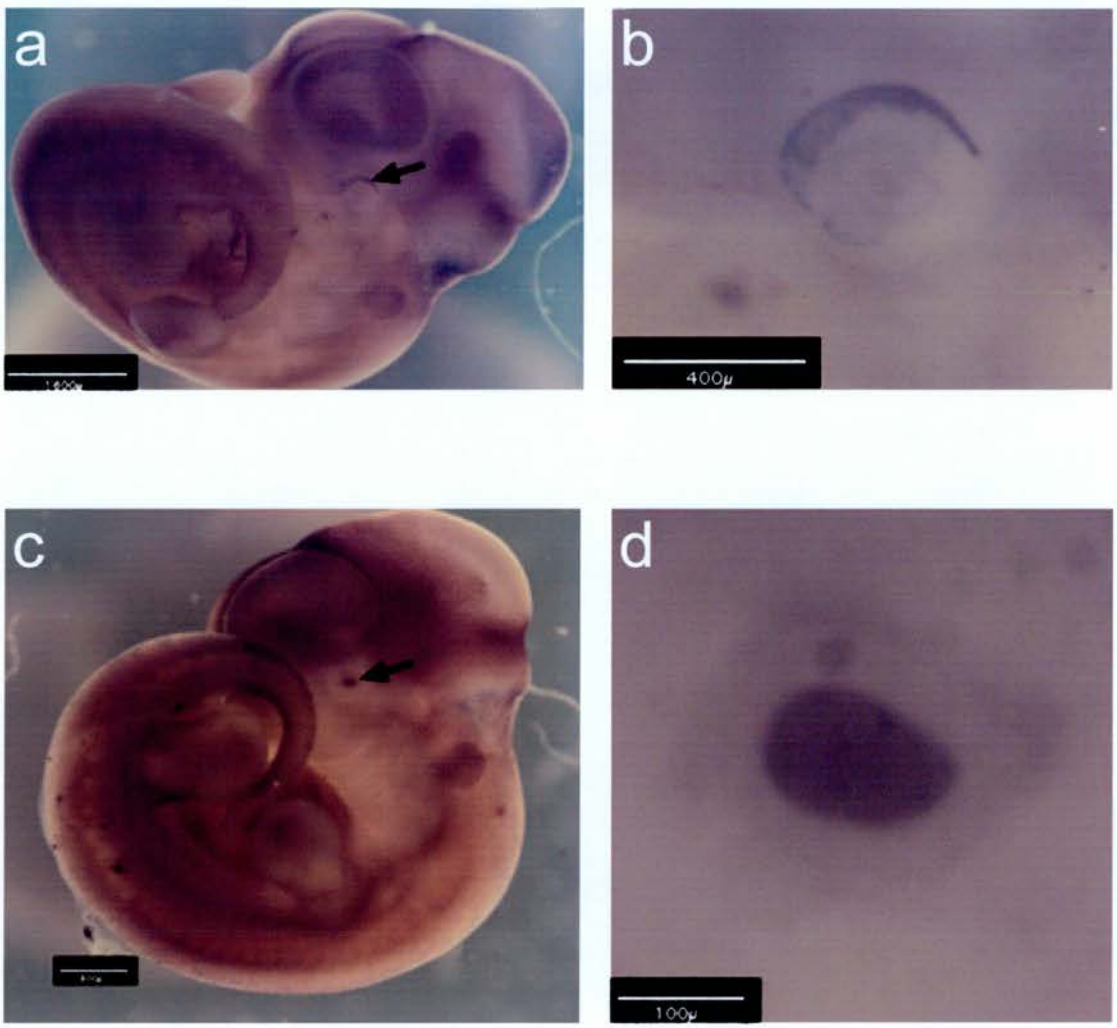


Figure 4.18 *Trale in situ* hybridisation at E11.5
In situ hybridisation of *Trale* shows sense expression to be widespread but strong signal can be seen in the retina (a, indicated by arrow). This can be seen more clearly at higher magnification, with the nasal/temporal axis from the left to the right side of the image (b). Antisense expression also seems to be widespread, but particularly strong in the lens (c, indicated by arrow). Lens expression is shown at higher magnification, with the nasal/temporal axis from left to right (d). Scale bars, 100 μm (d), 400 μm (b), 800 μm (c) and 1.6 mm (a).

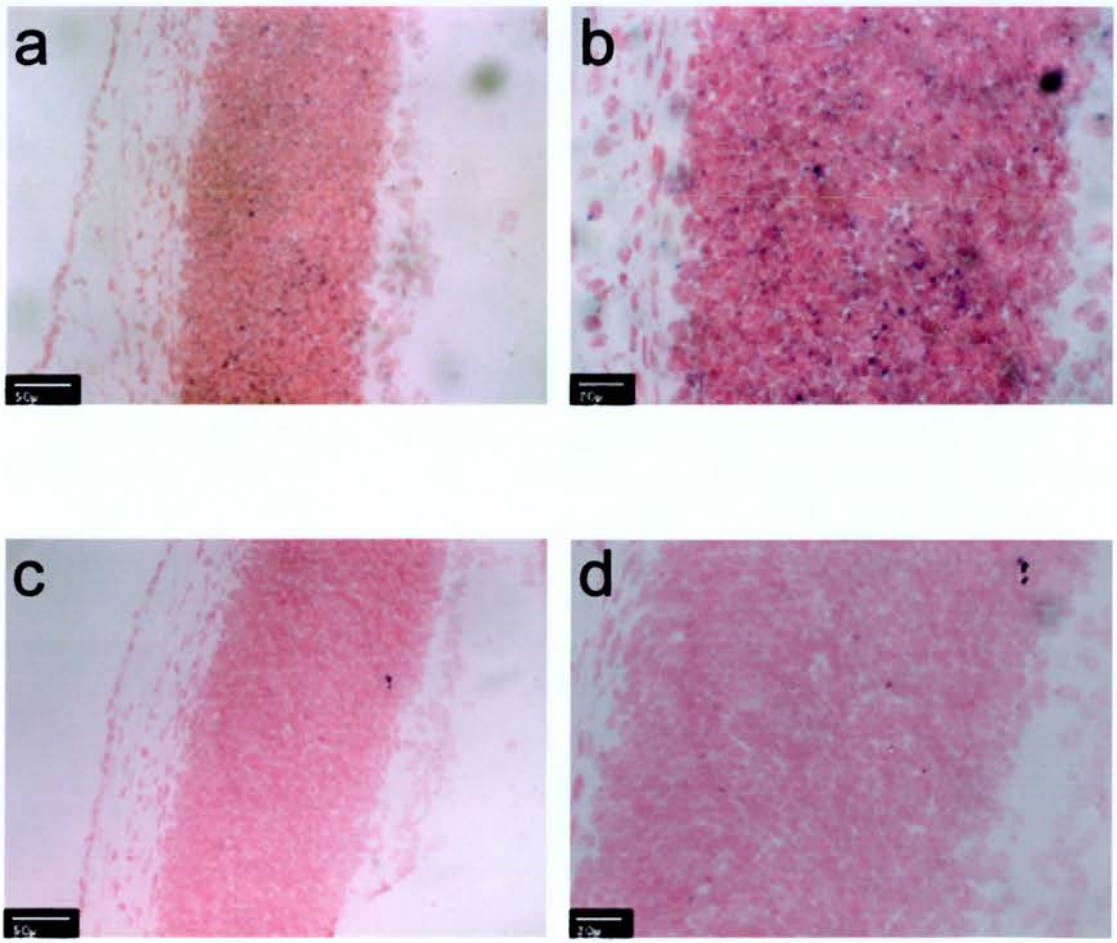


Figure 4.19 *Trale in situ* hybridisation in the neonate
In situ hybridisation on sagittal sections of the neonate (P1) eye shows signal in small, discrete points within the retina (a, experimental; c, control). Signal is also shown at higher magnification (b, experimental; d, control). All images show the choroid/lens axis from left to right. Scale bars, 20 μm (b, d); 50 μm (a, c).

Although signal seen in the neonate retina may not represent true expression, the early expression of *Trale* in the otic vesicle at E9.5 and E10.5 and its sense/antisense expression in the eye (retina and lens vesicle respectively) may be important in early development. The complementary domains of expression shown by the experimental and control probes could be indicative of co-regulation of gene expression. For example, expression from one strand in the lens may restrict expression of the opposing gene in the lens but not in the retina.

4.3.6 *Tset*

Expression of *Tset* is detected by RT-PCR (section 4.2) in the eye at E14.5 and E16.5 and in the neonate and adult. Expression was also seen in the brain, ovary and testis. Wholemount *in situ* hybridisations show expression of *Tset* at E10.5 (Figure 4.20) in the retina, but this is not observed at any of the other stages (E9.5, E11.5, E12.5, E14.5, E16.5, neonate or adult), despite previous indications by RT-PCR of eye expression at later stages of eye development as well as in the adult eye. There may also be brain expression. However, signal is also seen in the control, which may be indicative of background signal that is often associated with this organ.

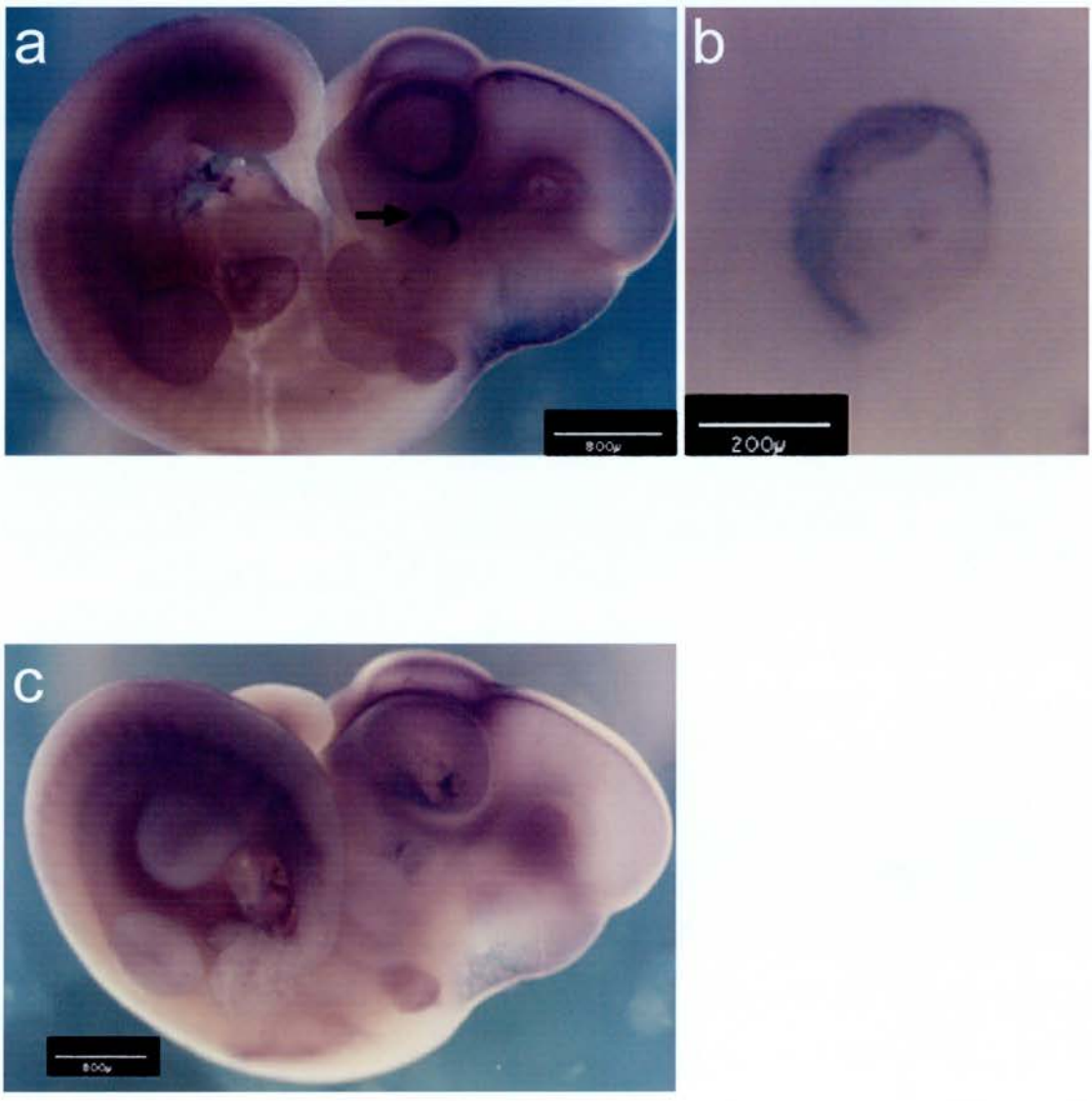


Figure 4.20 *Tset* wholemount *in situ* hybridisation
In situ hybridisation of *Tset* at E10.5 shows expression in the retina (a, experimental, indicated by arrow; c, control). Retinal expression can be seen more clearly at higher magnification, with the nasal to temporal axis from bottom left to top right (b). Scale bars, 200 μm (b); 800 μm (a, c).

The representative sequence of Mm.21485, BC017540, aligns to a gene encoding a novel protein with a nucleosome assembly protein (NAP) domain. This is true in mouse (NM_133745), rat (novel ENSRNOG00000000547) and human (NM_021648). These proteins belong to the testis specific Y encoded protein (TSPY) family. The relationship between NAP domain-containing proteins was assessed by constructing a phylogenetic tree (Figure 4.21), with sequences derived from Swiss-Prot (www.ebi.ac.uk/swissprot). It suggests that human TSPY is the most closely related to the novel NAP domain proteins identified here, followed by the SET translocation (SET) proteins. An alignment of the TSET, TSPY and SET proteins shows that the NAP domain region corresponds to the region of most homology between the proteins, with the N terminal end still showing similarity amongst the TSET proteins (Figure 4.22).

TSPY, the most closely related protein to *Tset*, was one of the first described genes on the human Y chromosome (Arnemann et al., 1991). As the name suggests, it has been shown to be specifically expressed in testis and testicular tumours (Arnemann et al., 1991; Zhang et al., 1992; Schnieders et al., 1996; Lau and Zhang, 2000; Dasari et al., 2001). Multiple copies exist in human (Vogel and Schmidtke, 1998) but it is a pseudogene in mice (Vogel et al., 1998) and its postulated role in spermatogenesis (Arnemann et al., 1991; Zhang et al., 1992) is yet to be confirmed. However, the domains of the TSPY/NAP/SET family suggests that it plays a role in gene regulation or in controlling replication or chromatin structure (Schnieders et al., 1996). As a member of this family, TSET may have a similar role in the developing eye.

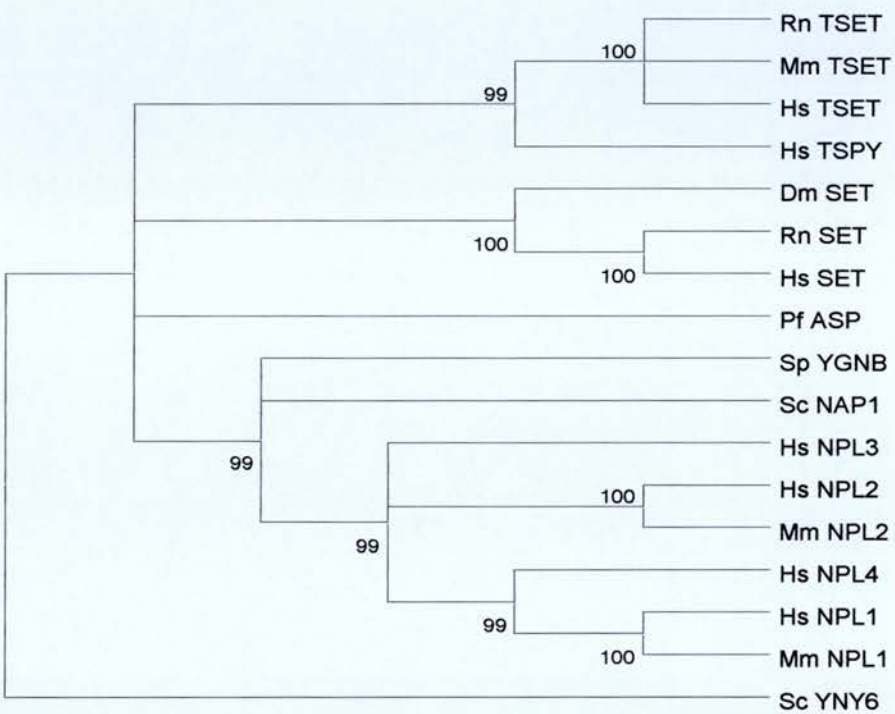


Figure 4.21 Phylogenetic tree of NAP proteins and TSET
TSET is more closely related to the human TSPY protein than to the other NAP domain-containing proteins. The neighbour-joining method was used to produce this tree and a cut-off value of 93 was applied to node confidence.

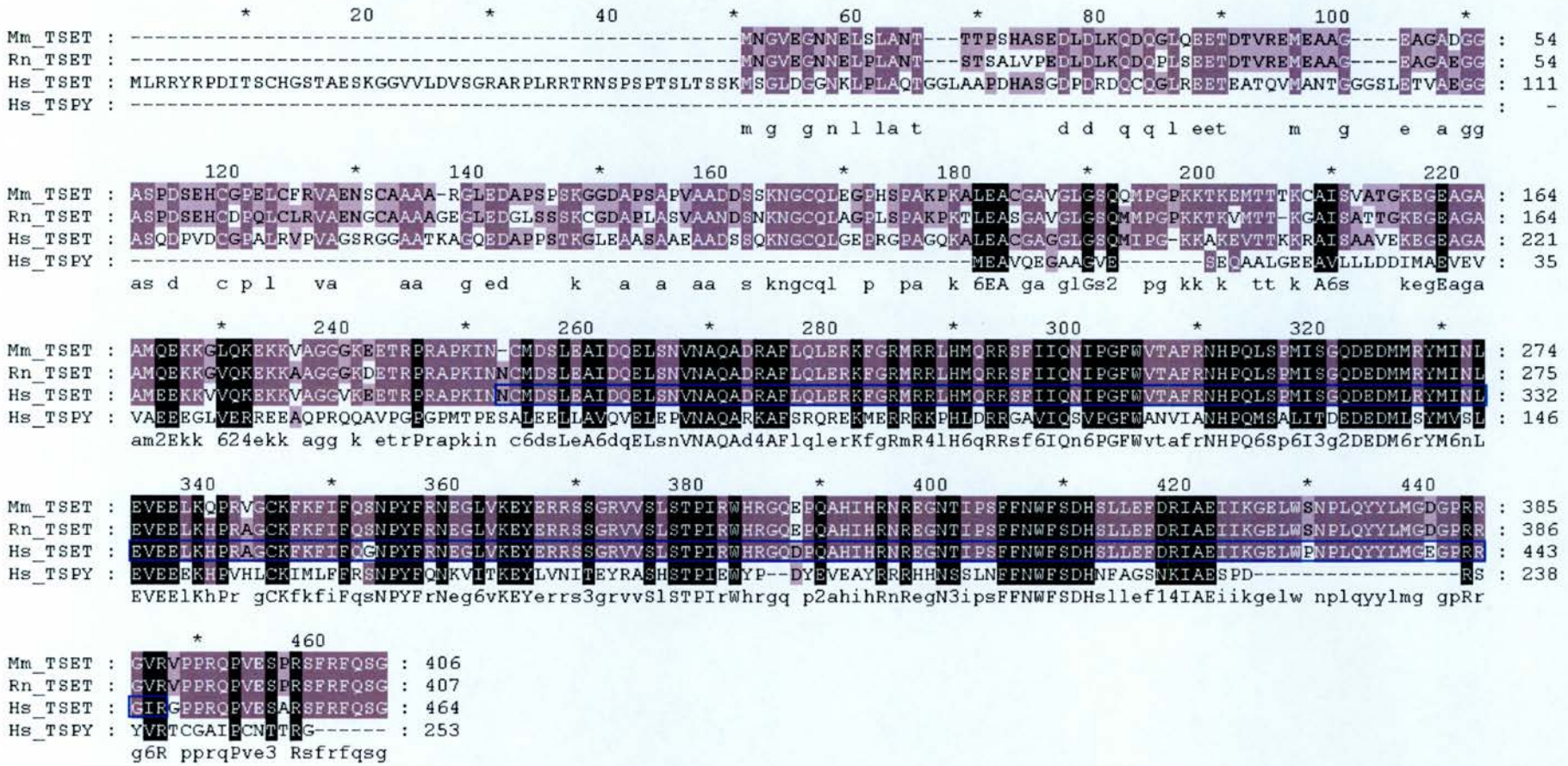


Figure 4.22 Alignment of the *Tset* proteins and human TSPY
Conservation is shown in the predicted NAP domain region (boxed in blue). Outwith this domain, conservation still exists between TSPY and the *Tset* proteins, as well as within the TSET proteins. Mm, *Mus musculus*; Rn, *Rattus norvegicus*; Hs, *Homo sapiens*.

4.3.7 *Bret*

Earlier RT-PCR results suggest that *Bret* is expressed at all tested stages in the eye (E14.5, E16.5, neonate and adult) as well as in the brain and testis (section 4.2). Using *in situ* hybridisation on cryosections, signal is seen throughout a central band in the neonate retina (Figure 4.23). This is similar to the pattern shown with *Trale*, however the signal is more intense, covers a more restricted region of the retina and each separate area of signal is larger. No expression was seen in the early embryo, neither was there any indication of signal at E14.5, E16.5 or in the adult, contrary to RT-PCR data.

In mouse, the representative sequence of Mm.38347, AK005160, aligns with all three exons of a gene (1500005I02Rik) with a BTB/POZ domain and a signal peptide sequence. In rat, the same is true (novel gene ENSRNOG00000003109), except for the presence of an additional, and seemingly untranslated, fourth exon. The fugu homologue, Q9DB72, also containing the BTB/POZ domain, aligns with the sequence. A Genscan prediction (AC103809.2.1.198764.8602.13713) and three ESTs are homologous to *Bret* in human. An alignment of these *Bret* predicted and novel proteins is shown (Figure 4.24). The predicted BTB/POZ domain is highly conserved across all proteins, but this conservation is maintained from this domain until the C terminus, with the exception of the fugu sequence, in which the final 96 residues are not included. A phylogenetic tree of the BRET proteins and BTB/POZ domain-containing proteins shows that BRET is most closely related to the mouse PPICAP protein and its human homologue LGALS3BP (Figure 4.25).

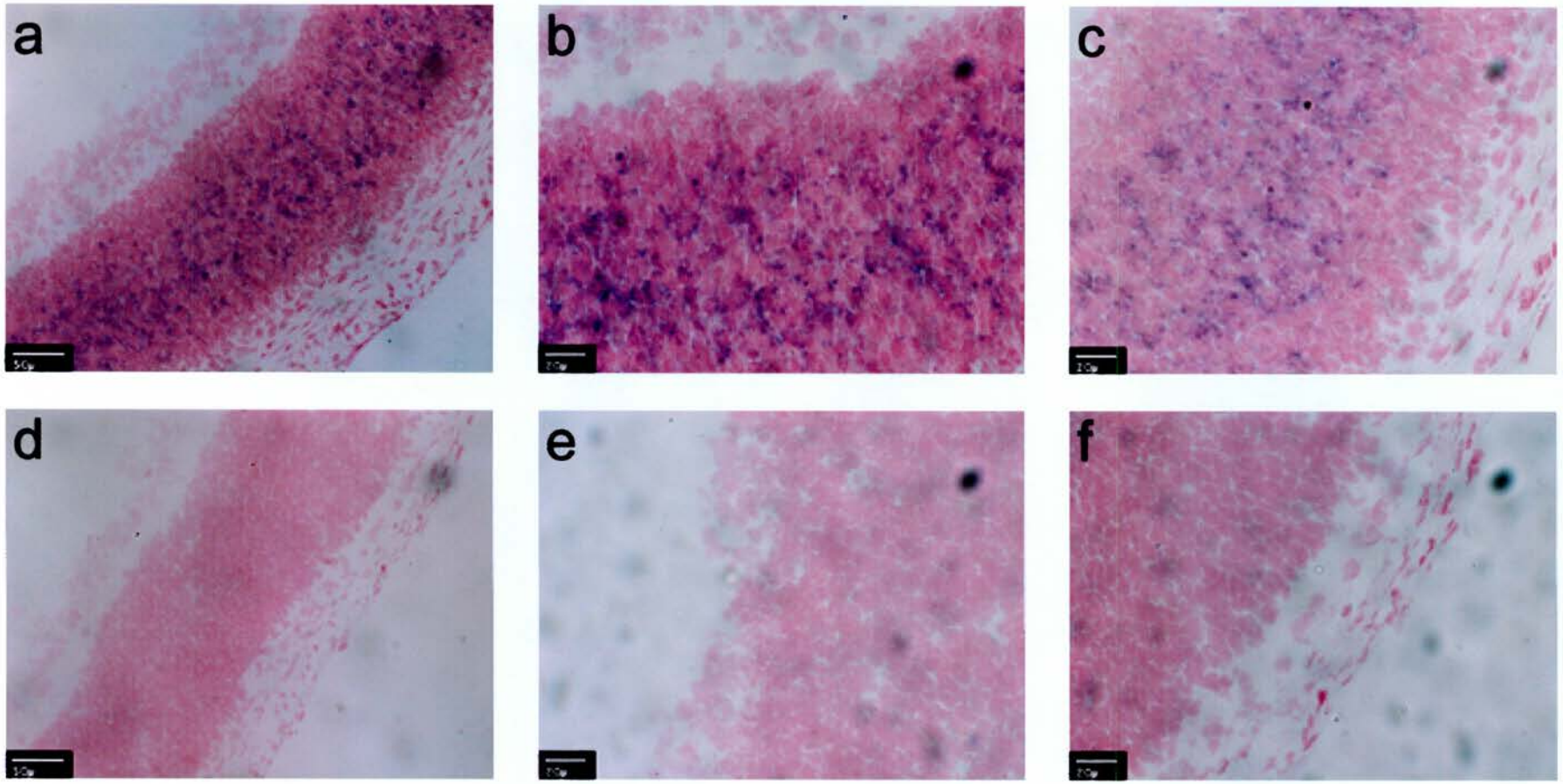


Figure 4.23 *Bret in situ* hybridisation
In situ hybridisation on sagittal sections of neonate (P1) retina show signal in small, discrete points in the neural retina (a, experimental; d, control). This signal seem to be absent from the inner (b, experimental; e, control) and outer (c, experimental; f, control) edges of the retina, although it is unclear whether these represent distinct layers of the neural retina. Images show the choroid/lens axis from bottom right to top left (a, c, d, e, f) and from bottom to top (b). Scale bars, 20 μm (b, c, e, f); 50 μm (a, d).

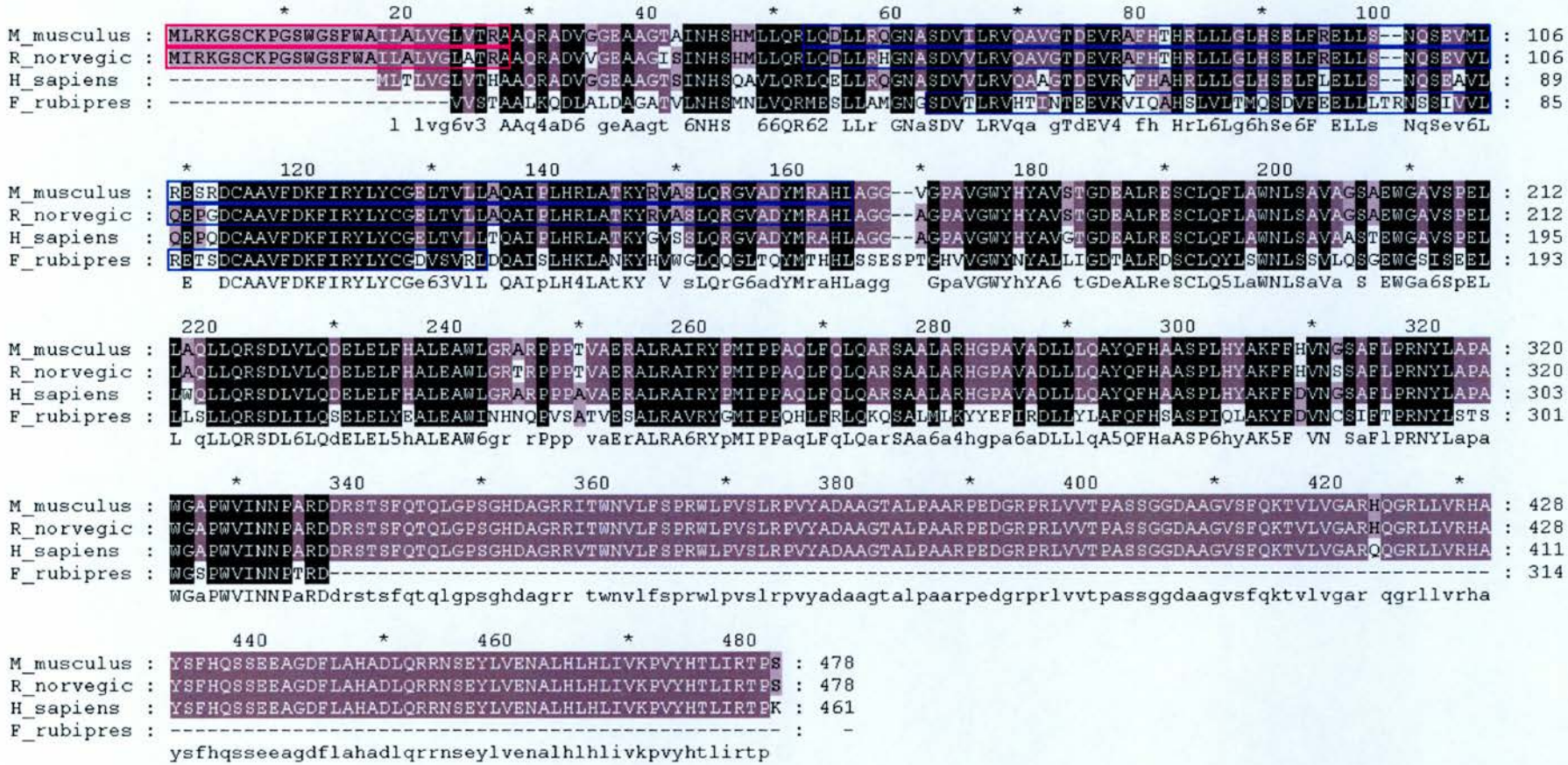


Figure 4.24 Alignment of the BRET proteins
A high level of conservation is shown throughout the predicted proteins. Predicted BTB/POZ domains are boxed in blue and predicted signal peptide sequences are boxed in pink. M_musculus, mouse; R_norvegic, rat; H_sapiens, human; F_rubripes, fugu.

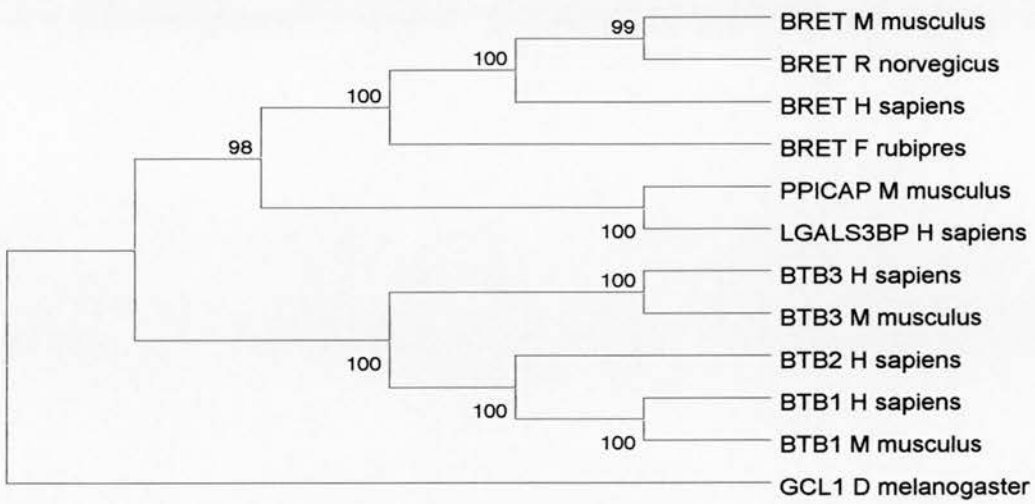


Figure 4.25 Phylogenetic tree of BTB/POZ proteins and BRET
BRET is more closely related to PPICAP and its human homologue LGALS3BP than to the other BTB/POZ domain-containing proteins. The neighbour-joining method was used to produce this tree and a cut-off value of 98 was applied to node confidence.

The high conservation of the BRET proteins across mouse, rat, human and fugu suggests an important conserved function across these species. The BTB/POZ domain was identified by two groups, hence the dual name encompassing BTB (Zollman et al., 1994) and POZ (Bardwell and Treisman, 1994). It has been shown to allow hetero- and homodimerisation of its member proteins (Bardwell and Treisman, 1994) and has been associated with transcriptional repression, through interaction with histone deacetylase co-repressor complexes (Deweindt et al., 1995; Wong and Privalsky, 1998; Huynh and Bardwell, 1998). The closely related PPICAP protein (peptidylprolyl isomerase C-associated protein) has been characterised as a cyclophilin C binding protein (Friedman et al., 1993). The physiological functions of both the cyclophilins and their binding proteins are as yet unknown. In human, LGALS3BP (lectin, galactose-binding, soluble, 3 binding protein) has been found to be upregulated in the serum of cancer patients (Iacobelli et al., 1986; Iacobelli et al., 1988; Iacobelli et al., 1993). Elevated protein levels were also observed in patients with AIDS and it was shown to stimulate host defense systems *in vitro*, prompting the authors to hypothesise a role in the immune system (Ullrich et al., 1994). The presence of a cysteine rich domain in both the human (Koths et al., 1993) and mouse (Friedman et al., 1993) proteins that is also found in MSR1 (macrophage scavenger receptor) supports this hypothesis. The BRET proteins, however, are not predicted to contain this domain, therefore it is less likely to have a closely related immune system function.

With respect to the *Bret* region, a locus associated with susceptibility to age-related maculopathy has been found in human at 17q25 in a genome scan that identified two significant associated loci (Weeks et al., 2001). However, this region is large (approximately 10.2Mb) and contains over 200 genes in the Ensembl genome assembly.

4.3.8 *Cln8l*

Using RT-PCR, *Cln8l* has shown expression in all tested stages of eye development (E14.5, E16.5, neonate and adult) and in adult brain (section 4.2). From wholemount *in situ* hybridisations, a low level of *Cln8l* expression can be seen in the optic cup at E9.5 and in the retina at E10.5 (Figure 4.26). The higher sensitivity of radioactive *in situ* hybridisation was used to examine more precisely the weak signal of *Cln8l* found in the eye at E10.5. Expression was found in the developing brain at E10.5 (Figure 4.27). This includes a region of expression extending from the telencephalon to the optic stalk. Although RT-PCR expression was seen at all stages, only E16.5 and adult *in situ* hybridisation sections produced visible signal. At these stages, expression is seen in the RPE (Figure 4.28). Morphology is poor in the adult eye due to sectioning procedures. No expression was seen at E11.5 or E12.5.

The representative sequence of Mm.23434 (AK005271) aligns with the entire five exon length of the novel mouse gene 1500016O10Rik, which has homologues in human (NM_031478), rat (novel ENSRNOG00000019914), zebrafish (novel ENSDARG00000015685) and fugu (SINFRUG00000136300). The rat gene is predicted to have a signal peptide sequence and the human and zebrafish genes are both predicted to have four transmembrane domains. Using a search against the SMART database (http://smart.embl-heidelberg.de/) for conserved protein domains, it was found that the mouse gene has a transmembrane domain and a TLC domain. An alignment of the genes from mouse, rat, human, fugu and zebrafish shows the position of these signal peptide, transmembrane and TLC domains (Figure 4.29). The TLC domain region is conserved across all genes. A phylogenetic tree was constructed to assess which of the known TLC domain-containing proteins *Cln8l* is more closely related to. The *Cln8l* proteins as well as other proteins with this domain are included, with sequences taken from Swiss-Prot (Figure 4.30). According to this, the CLN8 proteins are more closely related to *Cln8l* than the LAG and TRAM proteins.

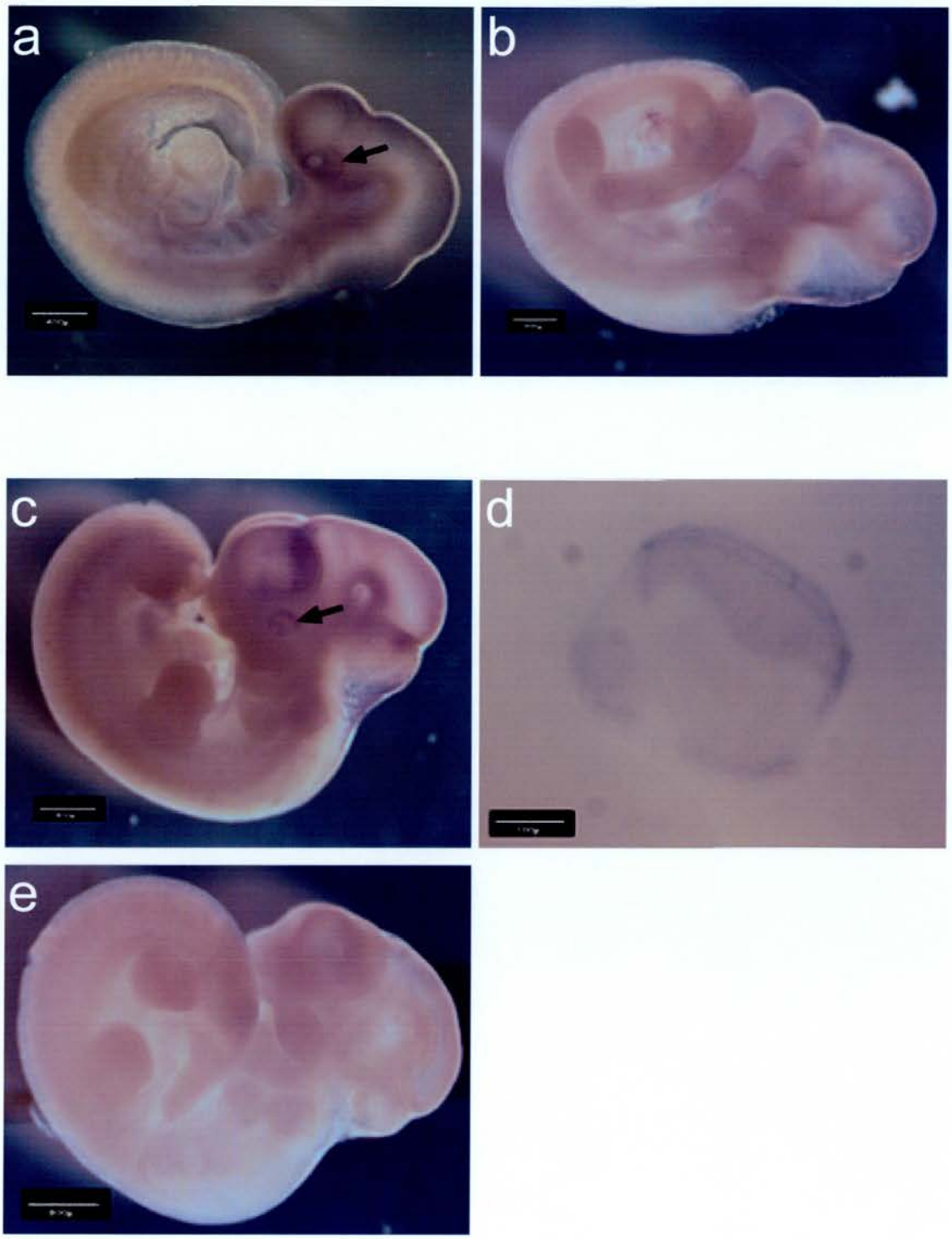


Figure 4.26 *Cln8l* wholemount *in situ* hybridisation
At E9.5 low signal can be seen in the optic vesicle (a, experimental, indicated by arrow; b, control). At E10.5, sense signal is also faint, but present in the retina (c, experimental, indicated by arrow; e, control). Retinal signal can be seen more clearly at higher magnification, with the nasal/temporal axis from left to right of the image (d). Scale bars, 100 μm (d); 400 μm (a, b); 800 μm (c, e).

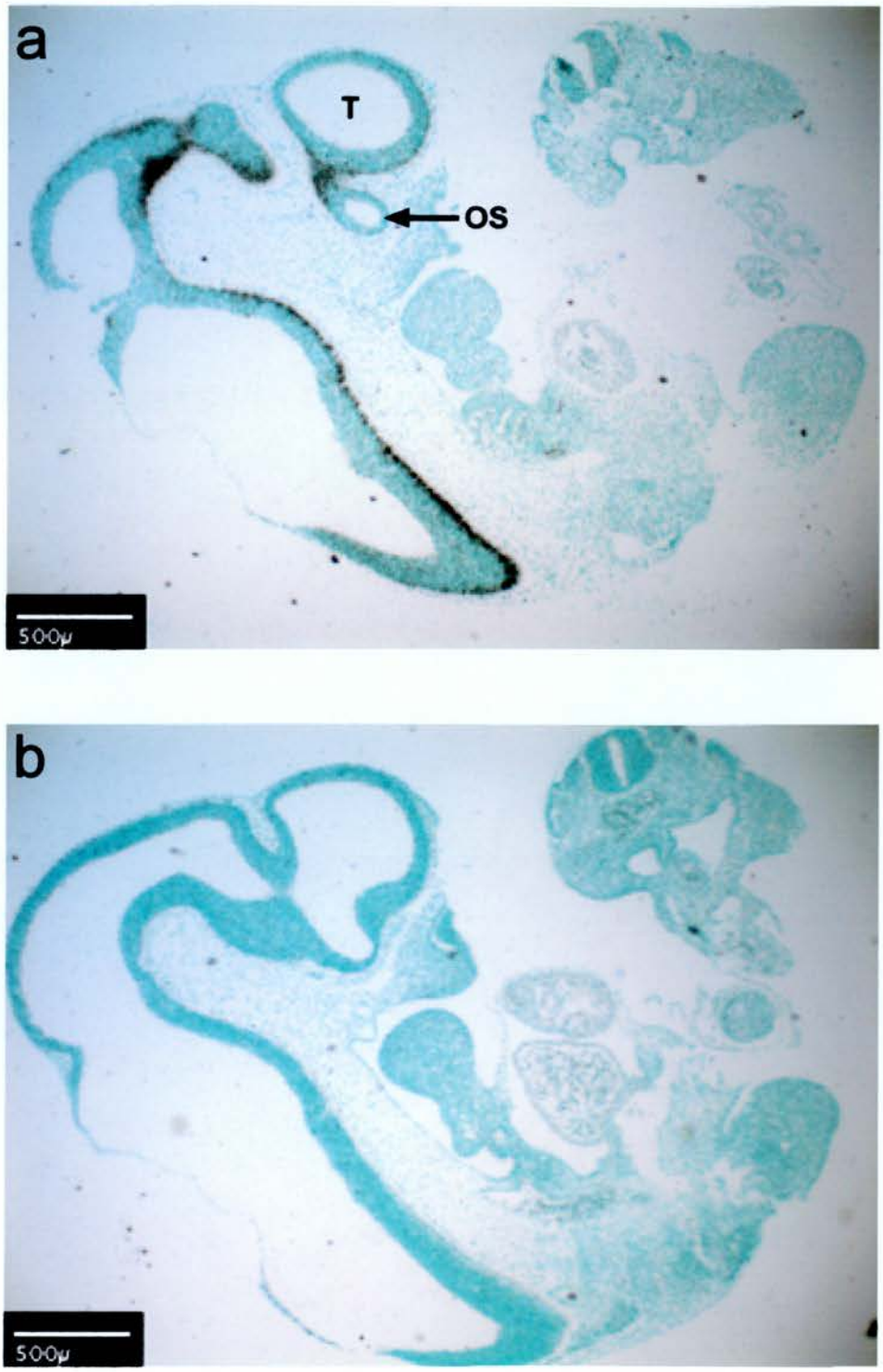


Figure 4.27 *Cln8l* radioactive *in situ* hybridisation
At E10.5 signal can be seen in the brain and in a region extending from the optic stalk to the telencephalon (a, experimental; b, control). T, telencephalon; OS, optic stalk. Scale bars, 500 μm.

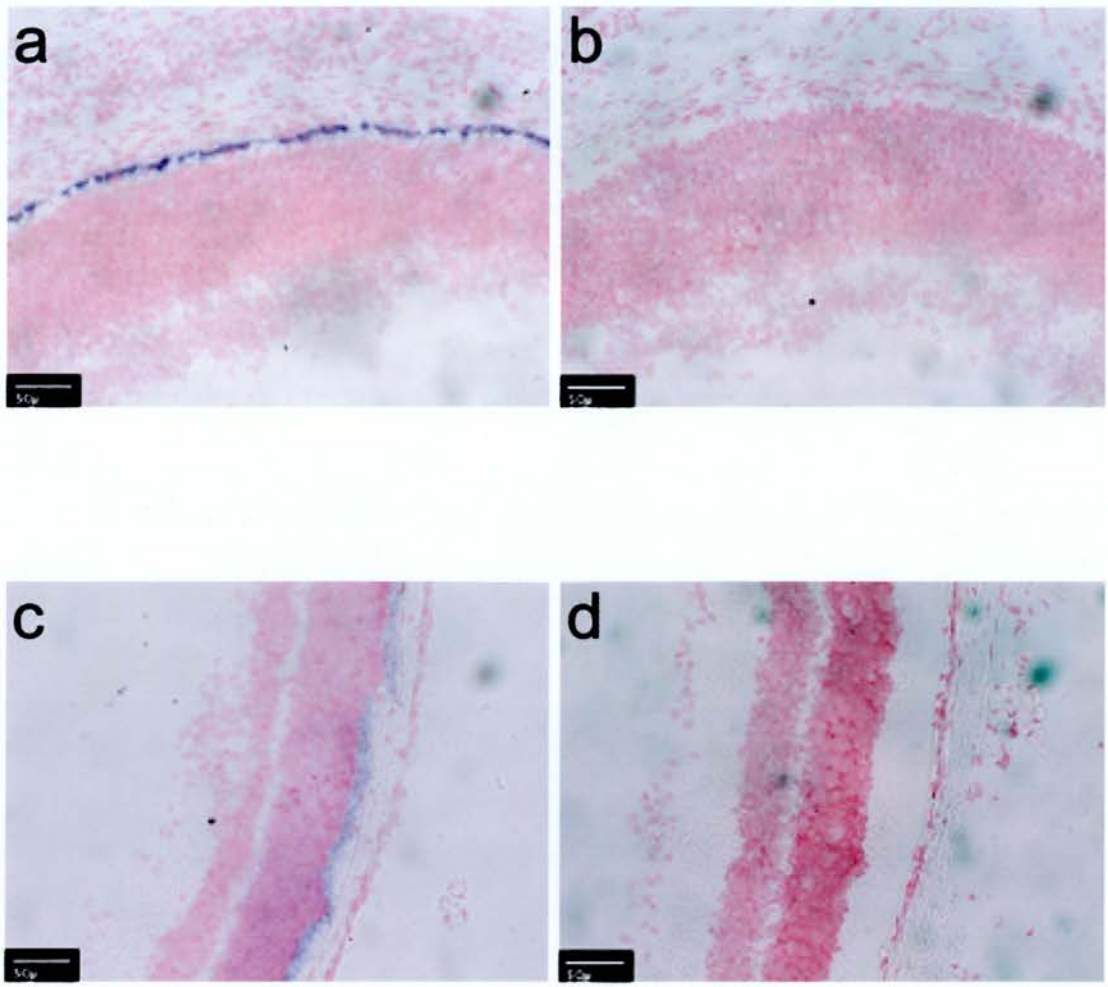


Figure 4.28 *Cln8l in situ* hybridisation
In situ hybridisation on sagittal sections of E16.5 embryo shows expression in the RPE, with the choroid/lens axis from top to bottom of the image (a, experimental; b, control). RPE expression is also seen in adult eye sagittal sections, with the choroid/lens axis from right to left of the image (c, experimental; d, control). Scale bars, 50 μm.

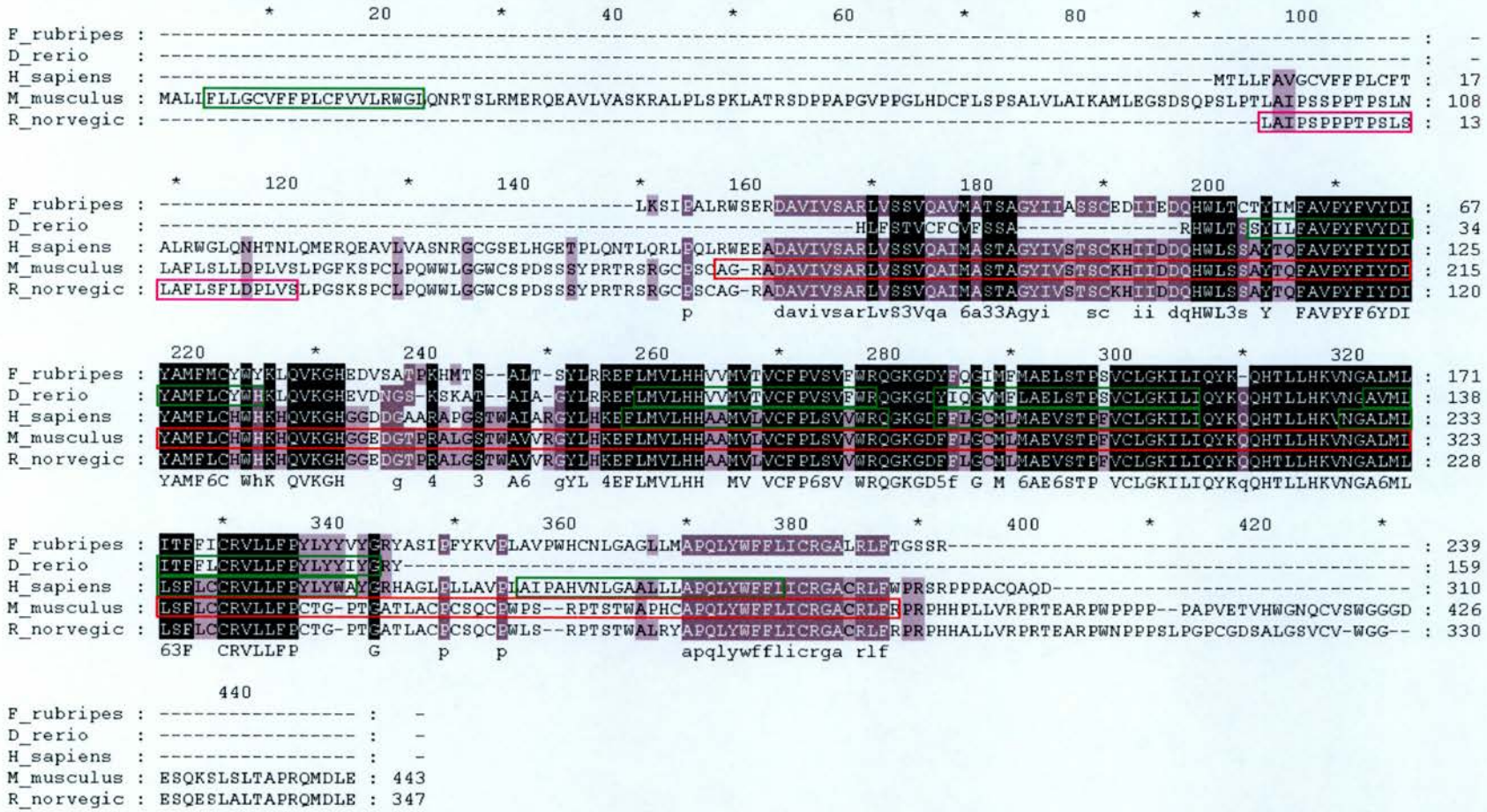


Figure 4.29 Alignment of the CLN8L proteins
Conservation is shown in the predicted TLC domain region (boxed in red). Predicted transmembrane domains are boxed in green and a predicted signal peptide sequence is boxed in pink. F_rubripes, fugu; D_rerio, zebrafish; H_sapiens, human; M_musculus, mouse; R_norvegic, rat.

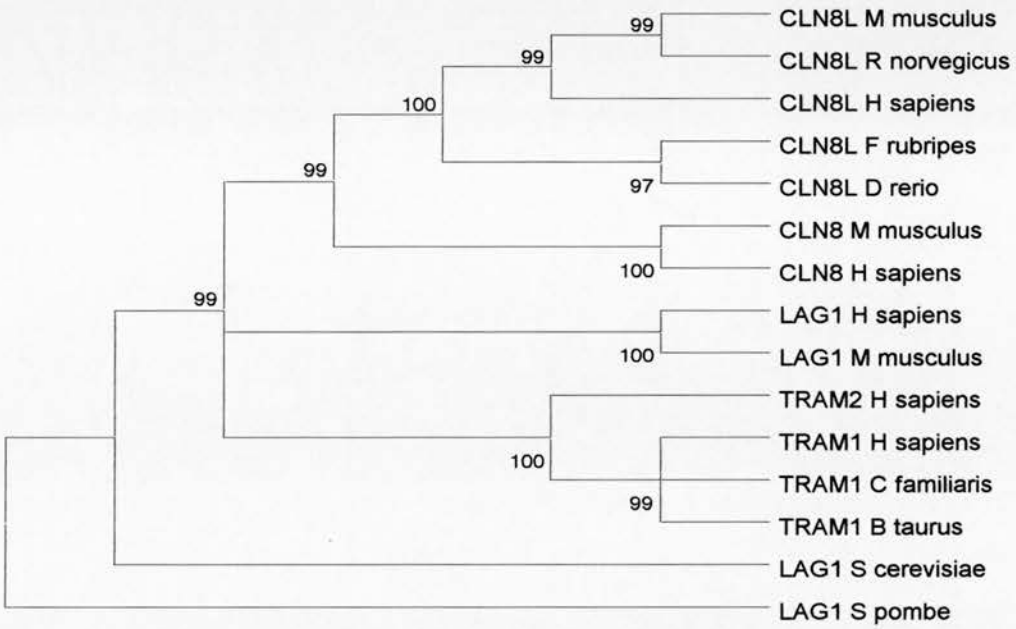


Figure 4.30 Phylogenetic tree of TLC proteins and CLN8L
CLN8L is more closely related to the CLN8 proteins than to the other TLC domain-containing proteins. The neighbour-joining method was used to produce this tree and a cut-off value of 97 was applied to node confidence.

The TLC domain contained within the CLN8L proteins is named after its commonality amongst the TRAM, Lag1p and CLN8 proteins (Winter and Ponting, 2002). The roles of these three types of protein have prompted the authors who gave the name to this domain to hypothesise roles for its members in lipid metabolism. Possible functions include activating lipid synthesis, protecting proteins from proteolysis, lipid transport or lipid sensing. Of the three types of protein that gave their name to this domain, CLN8 proteins seem to be most closely related to CLN8L proteins. *CLN8* has a missense mutation in patients with progressive epilepsy with mental retardation (EPMR), although it is in a non-conserved N terminal region and not in the TLC region (Ranta et al., 1999). In the same publication, the *mnd/mnd* mouse was reported to contain an insertion, within the TLC domain of *Cln8*, that is predicted to cause protein truncation. These mice display early-onset retinal degeneration and progressive adult-onset motor system degeneration, leading to premature death (Messer and Flaherty, 1986). Hence, it is not unreasonable to suggest that the function of CLN8L may be similar to other TLC proteins and that, as it is most closely related to CLN8, in addition to the fact that late expression of *Cln8l* is seen in the RPE, this protein may have an influence in maintenance of the retina. Signal in the brain and, more relevantly, between the telencephalon and the optic stalk, may suggest a possible role in the formation, function or maintenance of nerve fibres connecting the eye and the brain. However, the principal sensory association of the mature telencephalon is olfaction, with vision being the domain of the diencephalon. The expression data is not sufficient to indicate whether there are also connections between the optic stalk and the diencephalon at this stage. However, wholemount *in situ* hybridisation shows expression in the retina and, given later expression in the RPE, this could be expression in the presumptive RPE.

4.3.9 *Sapan*

In terms of RT-PCR data, *Sapan* shows expression in the eye at all tested stages (E14.5, E16.5, neonate and adult) as well as in the adult brain and ovary. At earlier stages of development, *Sapan* gives a striking expression pattern in wholemount *in situ* hybridisations. Signal is restricted to the developing eye, ear and brain and the dorsal neural tube at E9.5 (Figure 4.31), E10.5 (Figure 4.32) and E11.5 (Figure 4.33), but no expression is seen at E12.5. In the developing eye, expression appears to be in the optic vesicle at E9.5, in addition to the lens vesicle, which remains as an oval region without signal in cases where expression is restricted to the optic vesicle. At E10.5 and E11.5 signal is restricted to the lens vesicle. Sections of an E11.5 embryo that had undergone wholemount *in situ* hybridisation show expression in the central region of the lens vesicle (Figure 4.34). There is also signal in the cells lining the interior of the neural tube at this stage, known as the ependymal layer. Although RT-PCR data indicates that *Sapan* is expressed in the eye at all tested stages (E14.5, E16.5, neonate and adult), the *in situ* hybridisations carried out did not display specific signal at any of these stages.

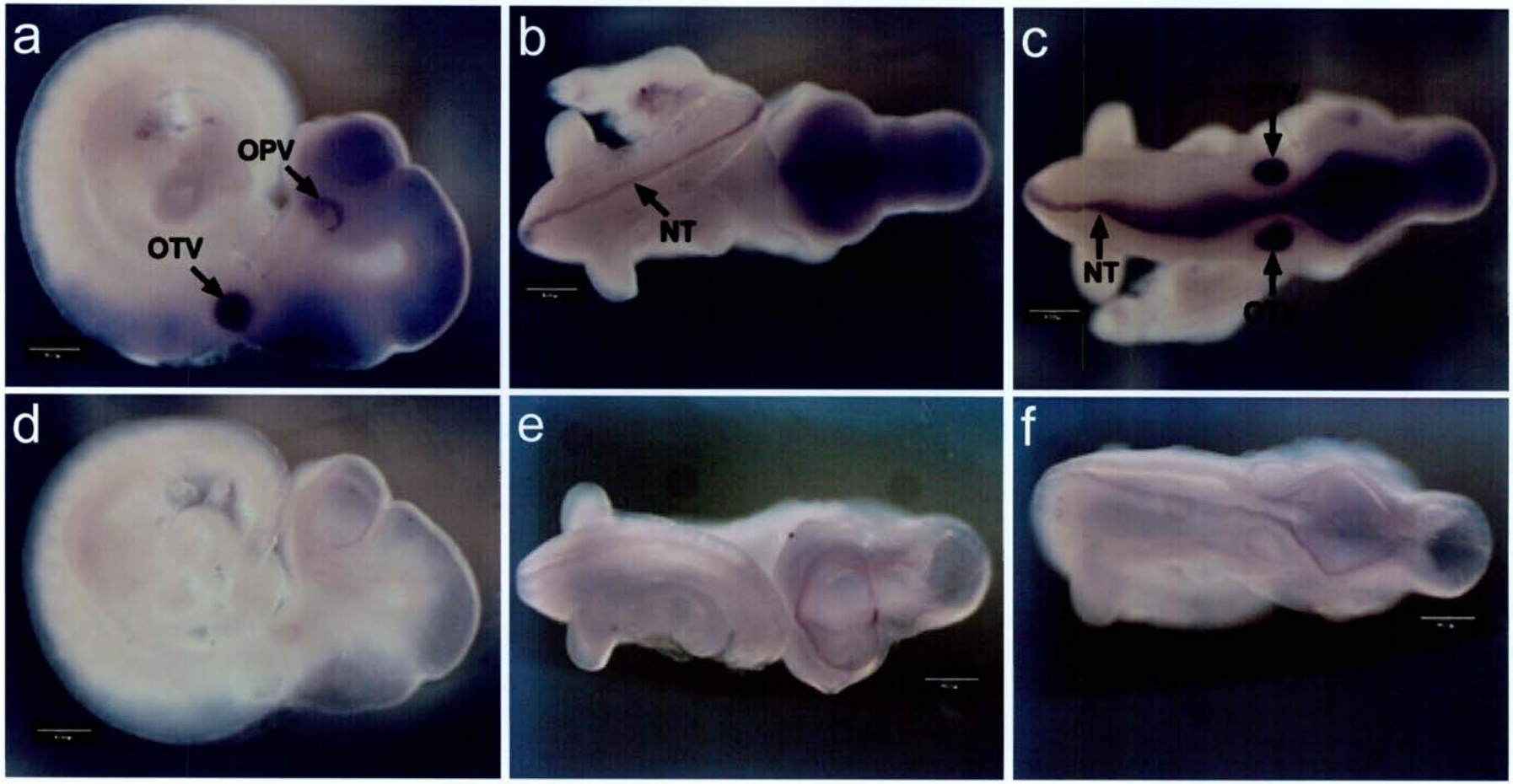


Figure 4.31 *Sapan* wholemount *in situ* hybridisation at E9.5
The lateral view of *in situ* hybridisation with *Sapan* shows expression in the otic and optic vesicles, developing brain and the neural tube (a, experimental; d, control). The ventral view shows expression in the developing neural tube more clearly (b, experimental; e, control). Otic vesicle and brain expression are seen in the dorsal view (c, experimental; f, control). OPV, optic vesicle; OTV, otic vesicle; NT, neural tube. Scale bars, 400 μm.

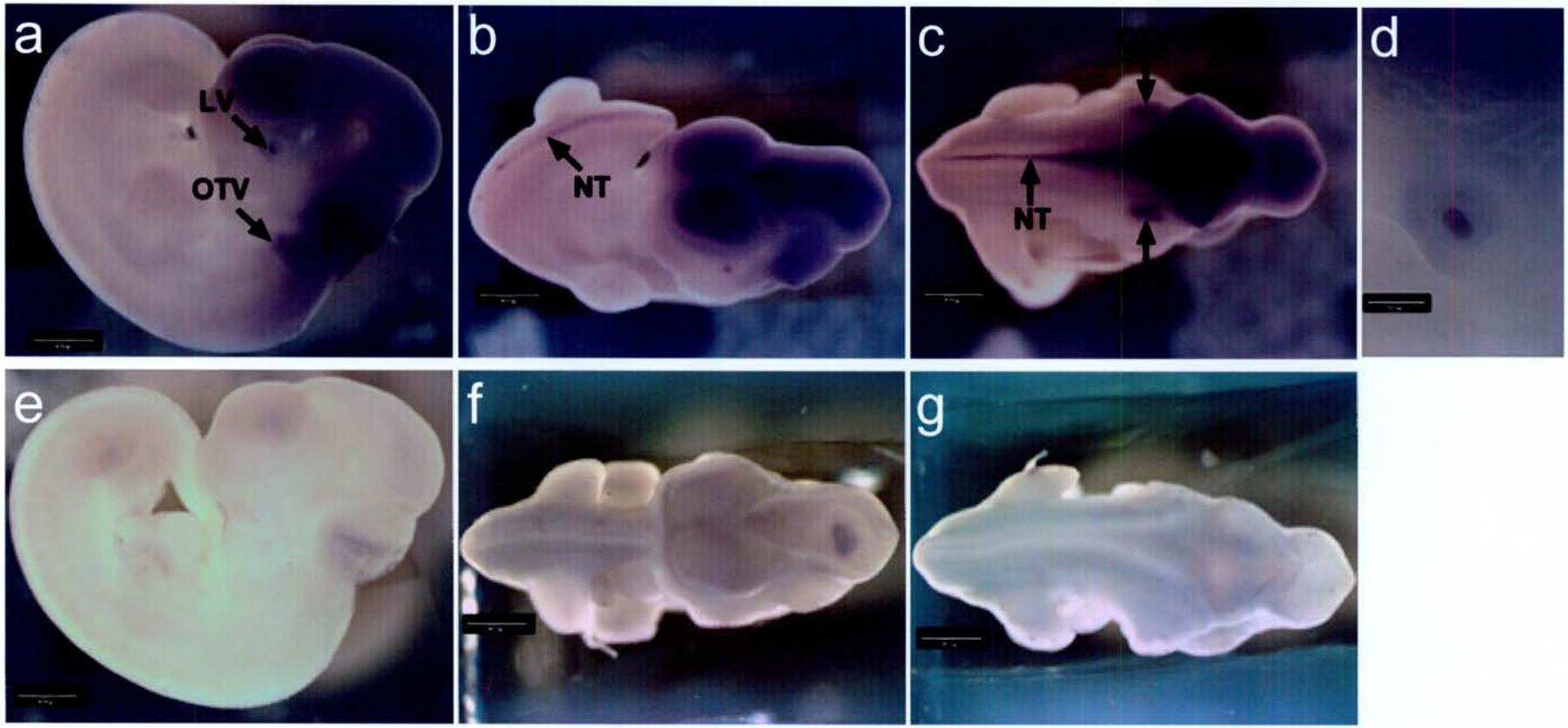


Figure 4.32 *Sapap* wholemount *in situ* hybridisation at E10.5
The lateral view of *in situ* hybridisation with Mm.44242 shows expression in the otic vesicles and the developing eye and brain (a, experimental; d, control). The ventral view shows a low level of expression in the developing neural tube (b, experimental; e, control). Otic vesicle and brain expression are seen in the dorsal view (c, experimental; f, control). A high magnification image, with the nasal/temporal axis from top left to bottom right, shows that eye expression is restricted to the lens vesicle at this stage (d). LV, lens vesicle; OTV, otic vesicle; NT, neural tube. Scale bars, 200 μm (d); 800 μm (a, b, c, e, f, g).

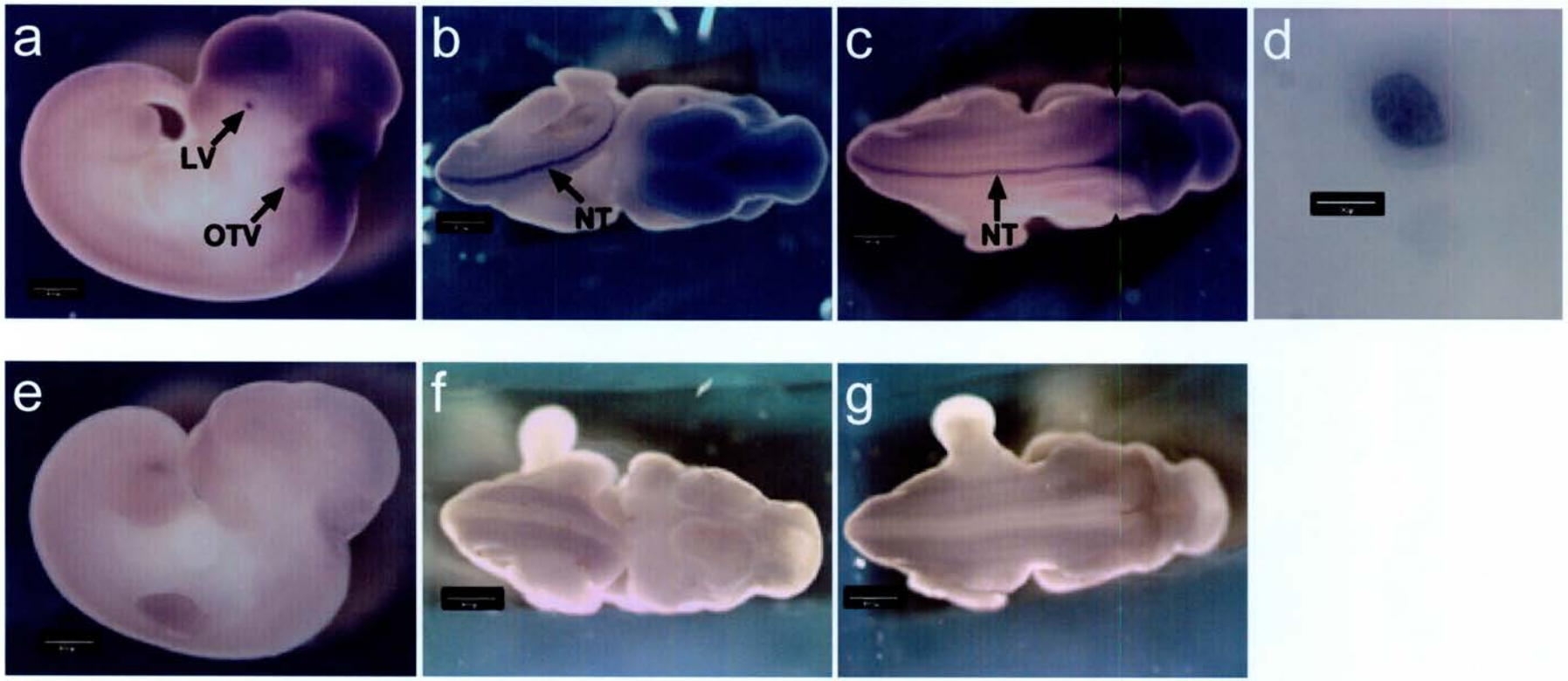


Figure 4.33 *Sapan* wholemount *in situ* hybridisation at E11.5
The lateral view of *in situ* hybridisation with *Sapan* shows expression in the developing eye and brain and a low level of expression in the otic vesicles (a, experimental; d, control). The ventral view shows expression in the developing neural tube (b, experimental; e, control). Brain expression and a low level of optic vesicle expression are seen in the dorsal view (c, experimental; f, control). A high magnification image, with the nasal/temporal axis from top left to bottom right, shows that eye expression is restricted to the lens vesicle (d). LV, lens vesicle; OTV, otic vesicle; NT, neural tube. Scale bars, 100 μm (d); 800 μm (a, b, c, e, f, g).

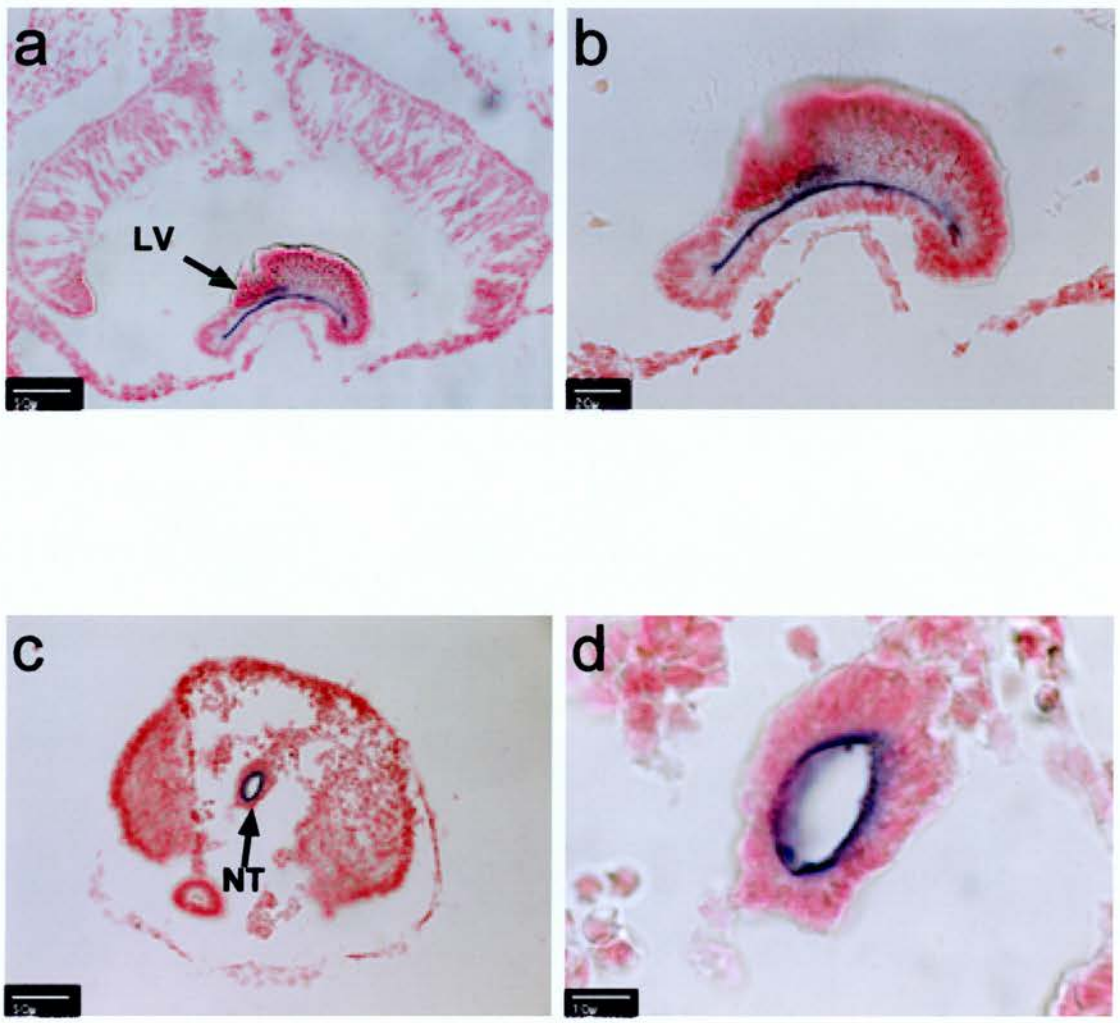


Figure 4.34 Sections of *Sapan* wholemount *in situ* hybridisation
E11.5 wholemount tansverse sections confirm that signal is present in the eye and in the developing neural tube. Eye signal is within the lens vesicle specifically, with the nasal/temporal axis from bottom to top (a; b, high magnification). Expression in the neural tube appears to be in the innermost cells, the ependymal layer, with the ventral/dorsal axis from top to bottom of the image (c; d, high magnification). LV, lens vesicle; NT, neural tube. Scale bars, 10 μm (d); 20 μm (b); 50 μm (a, c).

The representative sequence of Mm.44242, AK009408, aligns to a mouse protein containing two saposin A and two saposin B domains, encoded by a novel gene (ENSMUSG00000043430) that has homologues in both human (Q8N7T4) and rat (ENSRNOG00000006845). The only known proteins in mice to contain saposin domains are prosaposin (PSAP) and surfactant associated protein B (SFTPB). A phylogenetic tree was constructed to show the relationship of the novel proteins found here, using a saposin-related protein in fruitfly as an outlier (Figure 4.35). It shows that in terms of protein sequence, SAPAN is no more related to either of the two proteins. However, in terms of domain structure, the SAPAN proteins are more similar to PSAP (Figure 4.36). An alignment of these proteins, indicating the individual saposins that human prosaposin is cleaved into (Rorman et al., 1992) is shown (Figure 4.37). The region comprising the second saposin is largely absent in the SFTPB proteins, while there is homology between SAPAN and PSAP. Additionally, the regions outwith the saposins tend to be more similar to PSAP than to SFTPB, with the exception of the C terminal regions. This may imply that its function is more related to PSAP than to SFTPB. The novel gene aligning to the Mm.44242 representative sequence is located within an intron of, and in the opposite direction to, the *Sorcs2* gene in mouse, its provisionally named homologue, *SOR2*, in human and to an unannotated gene in rat, that is likely to be the *Sorcs2* orthologue.

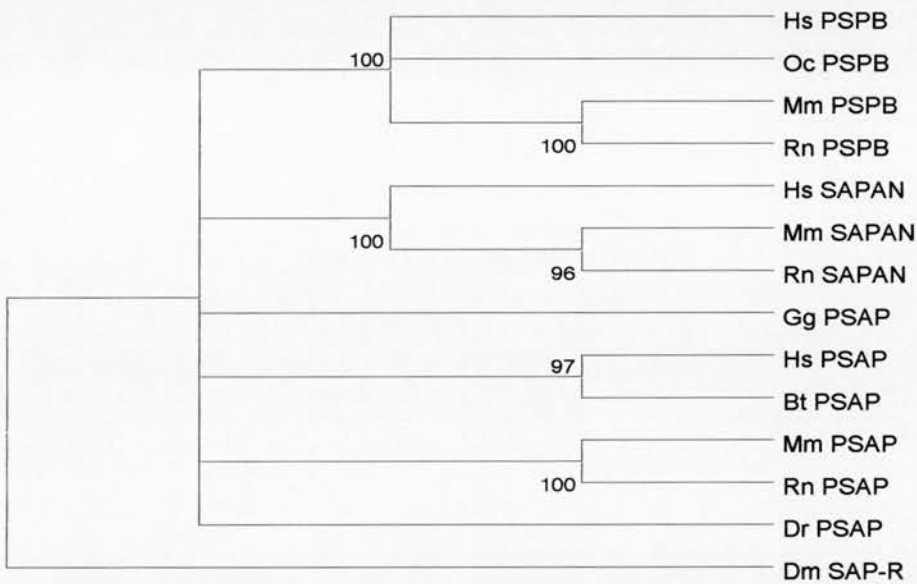


Figure 4.35 Phylogenetic tree of saposin domain proteins and SAPAN
SAPAN is no more related to the PSAP proteins than to the PSPB proteins. Additionally, the PSAP proteins appear to have diverged from each other at the same time as diverging from PSPB and the SAPAN proteins. The neighbour-joining method was used to produce this tree and a cut-off value of 95 was applied to node confidence.

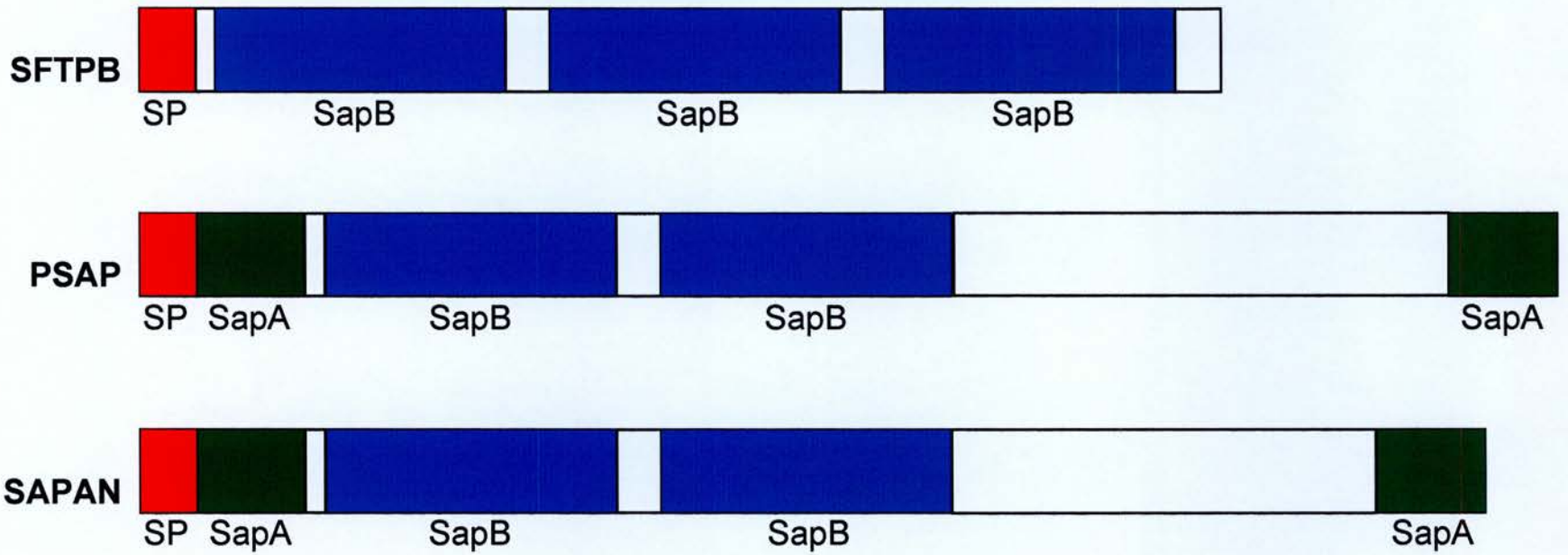


Figure 4.36 Domain structure of saposin domain proteins
Saposin domains are found in only two known mammalian proteins but are also present in SAPAN. The domain structure of SAPAN is similar to that of PSAP, with two SapA domains (green) in addition to two SapB domains (blue). However, SFTPB does not have SapA domains. All proteins have a signal peptide sequence (red). SFTPB, surfactant associated protein B; PSAP, prosaposin; SP, signal peptide; SapA (saposin A domain); SapB (saposin B domain).

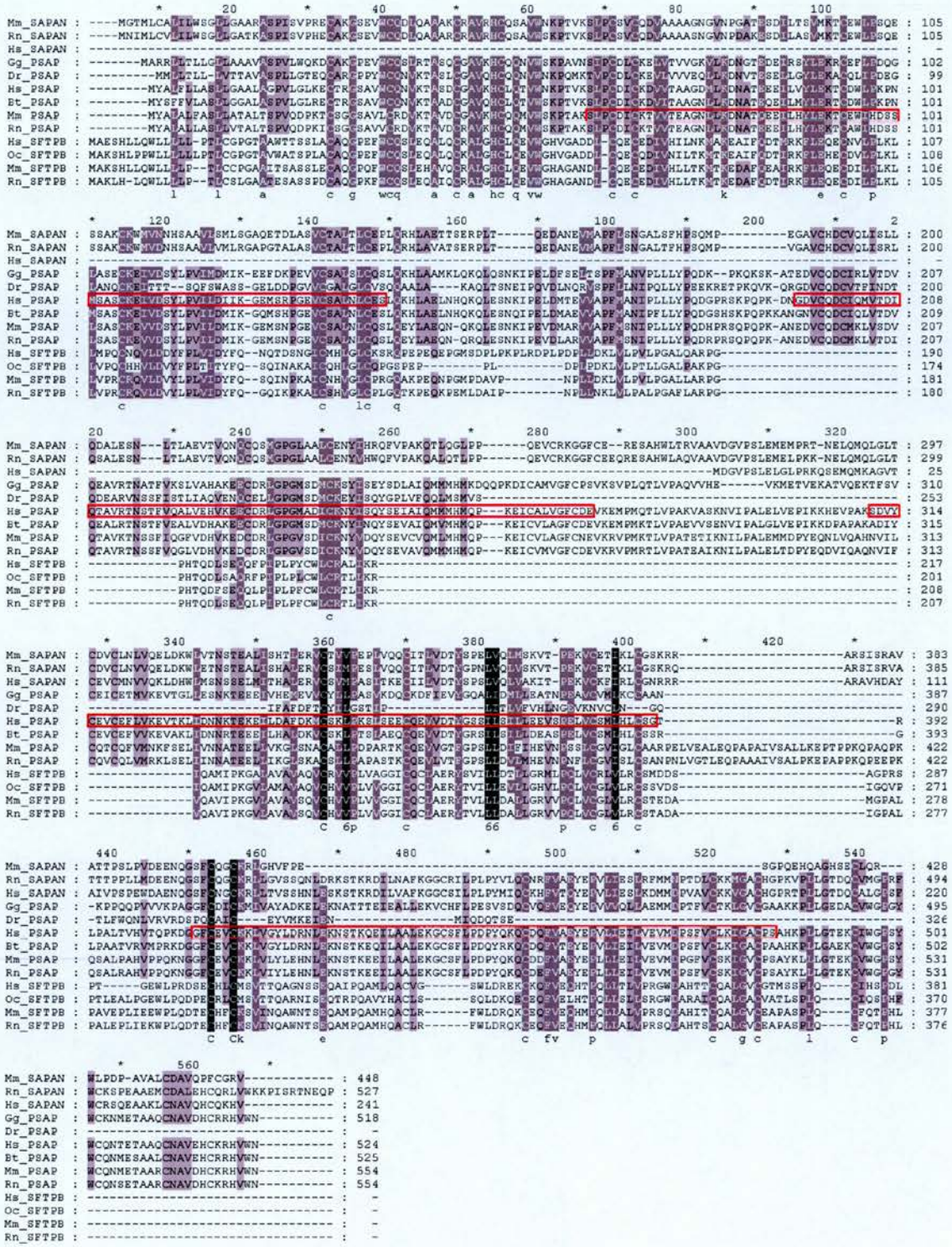


Figure 4.37 Alignment of SAP and SAPAN proteins

Areas boxed in red represent the individual saposins cleaved from the human prosaposin protein. Saposin C (3rd saposin) and the N terminal end of saposin D (4th saposin) show greatest conservation. More homology exists with PSAP in regions outside the saposins and also in saposin B (2nd saposin) which seems to be a largely missing region in SFTPB. Mm, *Mus musculus*; Rn, *Rattus norvegicus*; Hs, *Homo sapiens*; Gg, *Gallus gallus*; Dr, *Danio rerio*; Bt, *Bos taurus*; Oc, Oryctolagus cuniculus.

SORCS2 was recently identified as a member of the VPS10 domain-containing receptor family. It is expressed predominantly in the developing mouse brain, but also in the floor plate of the developing spinal cord, embryonic lung and heart and transiently in smooth and skeletal muscle and cartilaginous and connective tissue (Rezgaoui et al., 2001). *Sapan* is located antisense to *Sorcs2*, however, unlike *Tsd*, there is no expression in the antisense direction and it is located in an intronic region. This makes it unlikely that expression of either gene affects the other, but does not preclude the possibility. What might be the function of *Sapan*? Phylogeny suggests that *Sapan* is no more related to *Psap* than to *SFTPB* and that these genes diverged simultaneously. However, the protein structure of PSAP in terms of domain composition is the same as that predicted for SAPAN. *Psap* displays differential expression mediated by upstream regulatory regions (Sun et al., 2003), with high levels in the CNS, including the retina and ciliary body of the eye (Sun et al., 1994). The protein has a role in neurite outgrowth in mouse cells (Campana et al., 1996; O'Brien et al., 1994; O'Brien et al., 1995). PSAP has additional roles when it is proteolytically cleaved into four saposins, or sphingolipid activator proteins, that promote intracellular degradation of sphingolipids by specific hydrolases (O'Brien and Kishimoto, 1991). As a consequence, mutations in this gene are associated with glycosphingolipid storage diseases caused by its deficiency in humans (Paton et al., 1992; Hulkova et al., 2001; Rafi et al., 1993; Holtschmidt et al., 1991) and in its mouse homologue (Fujita et al., 1996; Matsuda et al., 2001). A patent (US6271196) filed by O'Brien in 2001 includes the DNA sequence corresponding to residues 318 to 339 of human PSAP. It describes a method for alleviating or preventing neuropathic pain by administration of an effective amount of this active fragment, although there is no evidence that these properties have been proved. *Sapan*, like *Psap*, is expressed in the CNS, however, in contrast to *Psap*, its eye expression is restricted to the lens and not the retina and ciliary body. Interestingly, a recent publication shows the similarity of J3-crystallin, one of three crystallins in cubomedusan jellyfish, to vertebrate saposins (Piatigorsky et al., 2001). As discussed in Chapter 1 (section 1.5.3), gene sharing in the lens is a common phenomenon, with crystallins carrying out different roles and being expressed at lower levels in other tissues (Piatigorsky et al., 1988; Piatigorsky and Wistow, 1989). These roles are

often as chaperone proteins or metabolic enzymes (Wistow and Piatigorsky, 1988; de Jong et al., 1989) and the roles of saposins mean that J3 crystallin fits the theory of enzyme recruitment for lens crystallins. It is not unreasonable to suggest that *Sapan*, showing differential expression in the developing mouse lens, may act as a crystallin. If this is the case, it is probable that it is a product of proteolytic cleavage, as there is evidence to suggest that J3 crystallin is cleaved from a precursor (Piatigorsky et al., 2001). *Sapan* expression in the brain, neural tube and the otic vesicle suggests it has a role in the development or function of these regions of the CNS. Expression in the otic vesicle is most prominent at E9.5, the stage at which the otic vesicle has just developed from the otic pit. Between E9.5 and E11.5, the otic vesicle changes shape and the endolymphatic appendage and semicircular canals, later involved in vestibular function, begin to develop. In the middle ear, the external pinna and Meckel's and Reichert's cartilage, from which the three auditory ossicles derive, have begun to develop. In spinal cord development, the dorsal half of the neural tube, known as the alar plate, eventually gives rise to the sensory region of the grey matter. The ependymal layer, in which expression of *Sapan* is seen at E11.5, is the innermost of three concentric layers that have differentiated by E10.5. The cells in this layer have a major role in the production and resorption of the cerebrospinal fluid. *Sapan* may therefore have roles in function or development in the eye, ear, brain and spinal cord, additional to eye and brain functions in the adult suggested by RT-PCR.

4.3.10 *Duf676*

RT-PCR results identified expression of *Duf676* in all tested stages of eye development (E14.5, E16.5, neonate and adult) and additionally in the brain, heart and testis. Although morphology is poor due to sectioning procedures, *in situ* hybridisation shows expression in either the choroid or the RPE in the adult eye (Figure 4.38). No *Duf676* expression was seen at early stages of embryo

development (E9.5, E10.5, E11.5 and E12.5), or at other stages at which RT-PCR results in the eye were positive (E14.5, E16.5 and neonate).

The representative sequence of Mm.87130 (AK019549) aligns with all eight exons of a mouse gene (4921533L14Rik) and a rat gene (ENSRNOG00000013587) with the DUF676 domain. These are also homologous to genes in human (NM_020819) and zebrafish (ENSDARG00000010247), with a longer but otherwise identical sequence to NM_020819 in human (Q9P2D6) and DUF676 domains as part of one alternative transcript in both organisms. An alignment of these predicted and novel proteins shows the high degree of homology in the domain region at the C termini of the proteins, however a much lower level of homology exists outside this domain (Figure 4.39).

The role of *Duf676* may be related to the function of the choroid or RPE cells in which it is expressed. In either case, *Duf676* may be associated with RPE maintenance and function. The novel DUF676 domain has wide taxonomical representation, which implies an important conserved function. However, the domain itself appears to be rare, with only 16 examples described in TrEMBL, and its function is unknown.

Duf676 lies within the region associated with North Carolina macular dystrophy (MCDR1) (Small et al., 1992). MCDR1 is a slowly progressive disorder with variable severity (Lefler et al., 1971; Small et al., 1992). However, the MCDR1 region is unrefined, spanning 24.7Mb from 6q14 to 6q16.2.

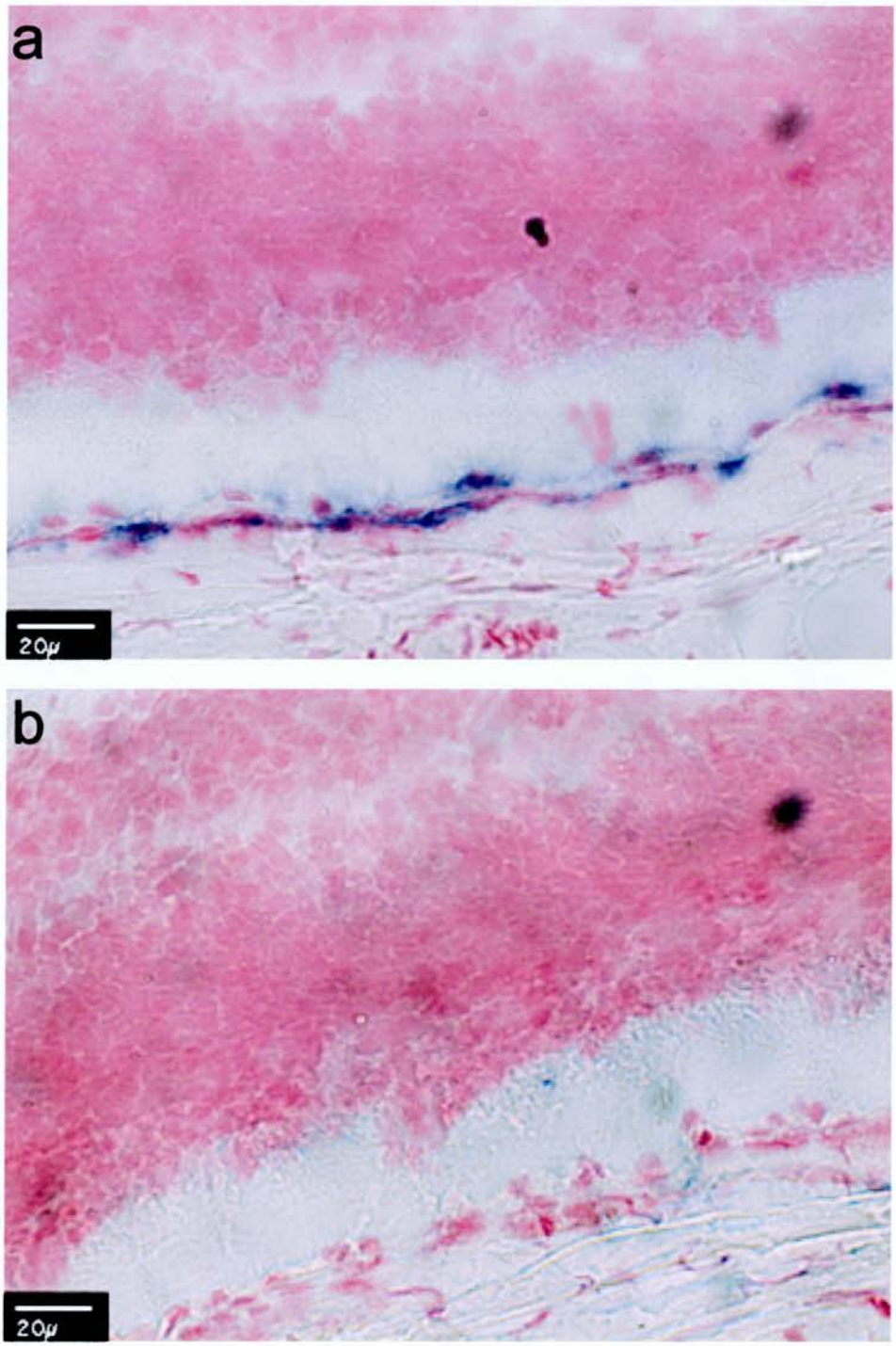


Figure 4.38 *Duf676 in situ* hybridisation
In situ hybridisation on sagittal sections of adult eye shows expression in either the RPE or choroid (a, experimental; b, control). Images show the periphery of the eye, with the choroid/lens axis from bottom to top (a, b). Scale bars, 20 μm.

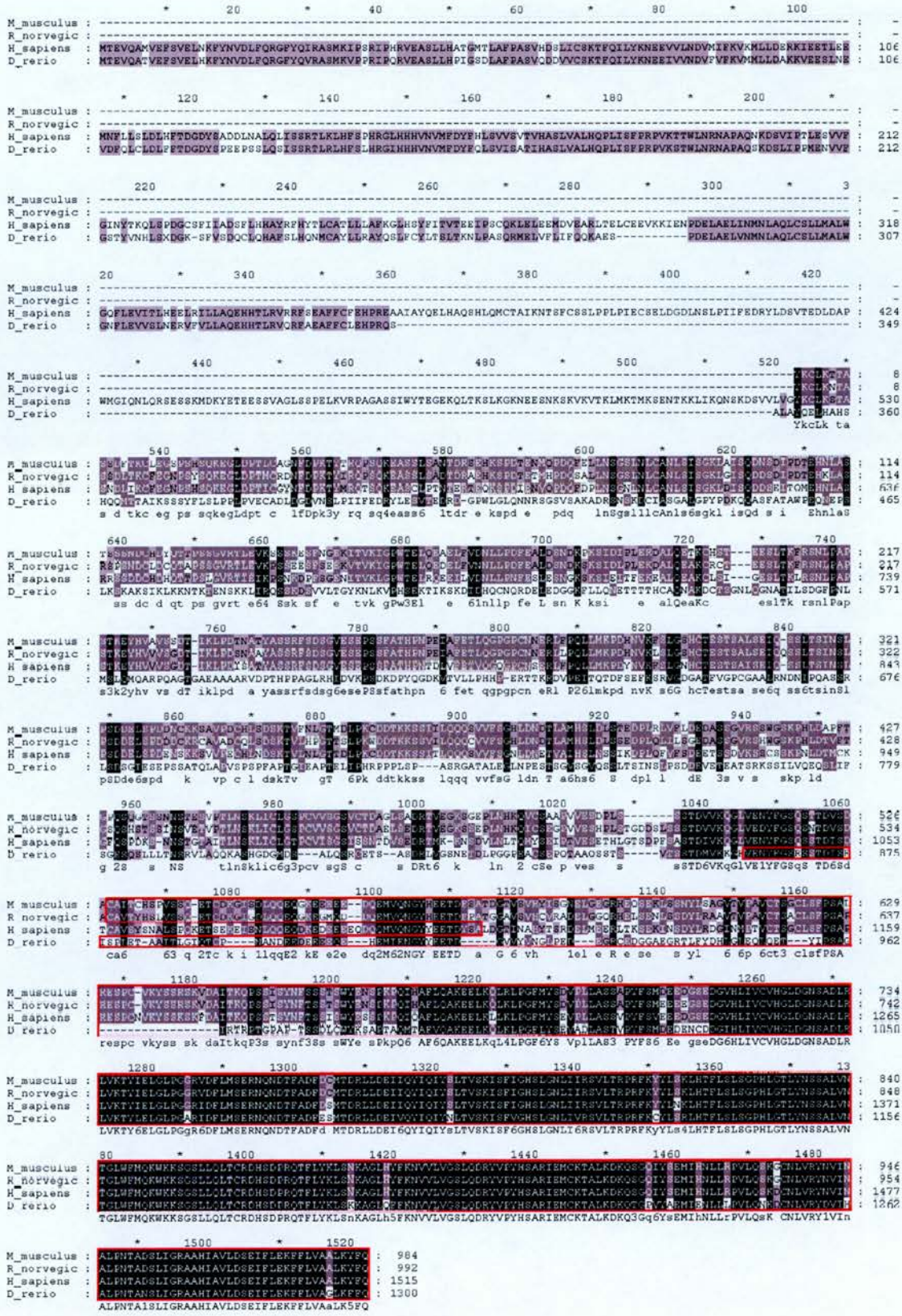


Figure 4.39 Alignment of the DUF676 proteins

The DUF676 domain region shows high conservation (boxed in red) and there is also conservation through the rest of the protein, although to a lesser extent. M_musculus, mouse; R_norvegic, rat; H_sapiens, human; D_rerio, zebrafish.

4.4 Discussion

Previously in this investigation, the success rate of DDD in producing eye relevant genes has been deduced to lie between 45% and 50% based on known genes. If this is the case for novel clusters as well as those representing known genes, 14 to 16 of the 31 genes that were chosen for further expression analysis would be expected to be eye relevant. As 15 of these clusters are preferentially expressed in the eye, and assuming this is an indication of their relevance in the development and function of the eye specifically, it would appear that this is indeed the case. However, this assumes that non-preferential eye expression is indicative of no specific eye relevance and this may not always be the case. Nevertheless, 50% seems to be a good approximate success rate for this method on the basis of RT-PCR data.

Amongst the clusters displaying preferential expression in the eye, all except one of those showing transcription in another tissue exhibited brain expression. This is not unusual, as the retina can be thought of as an extension of the brain, with its development ensuing from the neuroepithelial layer of the cephalic neural folds, the neuronal cell types present in the retina and its obvious sensory function. *Otx2*, for example, is important in early head development and is therefore important in brain and eye function (Matsuo et al., 1995). Perhaps more surprising is the number of preferentially expressed clusters that display ovary and, particularly, testis expression. As all RNA samples were quantified, and a GAPD control was employed to ensure no bias in expression, this appears to be a real phenomenon, however there is no obvious reason for this to be the case.

Five of the 11 novel clusters were negative in wholemount *in situ* experiments. This may have been due to a combination of factors. It is possible that some of these clusters have derived from ESTs that are the result of spurious transcription, resulting in little or no expression and certainly no expression detectable by this method. There may also be functional clusters that have low expression that can not be detected in this way. A third alternative is that the probe designed was ineffective in

these hybridisation experiments. However, *Dct* represents a good positive control, with expression shown in the RPE. The *Otx2* and *Crygf in situ* hybridisations, showing expression in the RPE and lens respectively, validate the methods used for the design and production of probes for the 11 novel clusters.

Of the 11 novel clusters with which preferential eye expression at E14.5, E16.5, neonate or adult stages was detected by RT-PCR, only five have been confirmed by *in situ* hybridisation. The use of *Dct* as a positive control suggests it is unlikely for *in situ* experiments to have failed due to experimental procedures during *in situ* hybridisation. However, whilst the known clusters representing *Aipl1*, *Crygf* and *Otx2* were used as controls for experimental procedure using the same probe synthesis techniques during the wholemount *in situ* hybridisations, this was not done with *in situ* hybridisations carried out on sections. Therefore, the nature of individual riboprobes may have had an effect or, alternatively, if these clusters are expressed at low levels, expression detected by RT-PCR may not be detectable using the *in situ* hybridisation technique employed. This implies that although the success rate of between 45% and 50% for eye specificity amongst known genes seems to apply to RT-PCR data of novel genes, this does not necessarily translate to detectable *in situ* expression. As *in situ* techniques are less sensitive than RT-PCR, discrepancies between RT-PCR and *in situ* results in this study may be caused by this difference. Although this is possible, for many of the clusters some positive RT-PCR results were confirmed by *in situ* analysis whilst others were not. As RT-PCR was not quantitative, there may be an underlying difference in levels of expression that *in situ* is sensitive to. This could be addressed by using quantitative RT-PCR techniques.

The E14.5 and E16.5 stages are under-represented amongst the results, with no clusters expressing at E14.5 and just one cluster showing expression at E16.5. This could be a consequence of lower expression at these stages or the experimental procedure used. However, the *Dct* control did not appear less convincing at these stages than at the neonate stage, which is the stage that has shown expression with most clusters and the procedure itself was not changed.

Surprisingly, two of the six early expressing novel clusters display sense and antisense transcription (*Tsd* and *Trale*). Recent studies have found 2,481 and 1,600 pairs of sense/antisense transcripts in mouse (Kiyosawa et al., 2003) and human (Yelin et al., 2003), respectively. This indicates that at least 17% of the mouse and 8% of the human genes have antisense transcripts. The true number may be higher, as the non-coding nature of many antisense transcripts result in difficulty in their discovery when using bioinformatic methods. Many antisense transcripts are discovered experimentally through investigation of the sense RNA. Such examples include developmentally important genes such as *Hoxd3* and antisense *Dxoh3* in mouse (Bedford et al., 1995), *FGF2* and antisense *gfg* in Xenopus (Kimelman and Kirschner, 1989), rat (Murphy and Knee, 1994; Li and Murphy, 2000) and human (Murphy and Knee, 1994; Knee et al., 1994) and *Msx1* and its antisense transcript (Blin-Wakkach et al., 2001). In both cases presented in this thesis, expression is seen in the eye in both sense and antisense directions, with one gene showing expression within the same domain of the eye in both directions. It has recently been reported that sense and antisense transcripts tend to be isolated from the same cDNA libraries (Kiyosawa et al., 2003). Although this is not proof that the transcripts will be expressed in the same cell, this coincident library expression is associated more with antisense transcripts than with bi-directional transcripts that are antisense to the intronic region of another gene. The antisense expression seen with *Tsd* and *Trale* could be verified by using specific amplification of the antisense sequence by RT-PCR, as a positive result would show that transcription is taking place in the antisense direction.

For Mm.150838 and *Duf676*, there is uncertainty about the identity of positive cells in the periphery of eye sections. These could be choroidal or RPE cells as the nature of the sections make it impossible to be certain. Better quality sections may resolve this ambiguity. Colocalisation with a specific antibody to either of these cell types or *in situ* hybridisation with two probes, one for the cluster and one for a gene known to be specifically expressed in one or the other of these cell types, could definitively identify the cell types expressed by Mm.150838 and *Duf676*. In addition, with the exception of Mm.150838, all clusters tested by *in situ* hybridisation were shown to

be expressed in the adult brain by RT-PCR. It would therefore be useful to carry out *in situ* analysis on adult brain in order to identify expressing cell types which may then be related to cells in the eye expressing these sequences.

Of the 11 novel clusters that have confirmed preferential eye expression, nine display preferential expression by *in situ* hybridisation. Domains of expression and bioinformatic investigation have allowed the postulation of a variety of roles for these novel proteins: controlling expression of opposite strand transcripts; determining cell fates by their positional identity in the retina; developing and maintaining neuronal connections between the retina and the brain; maintaining lens structure and function; development and function of the CNS; regulation of gene expression or mitosis; lipid metabolism; RPE maintenance; horizontal cell proliferation and/or maturation; retinal development and/or differentiation. These represent interesting novel genes that may be important in eye development and function. Further studies of these genes and their homologues may lead to a greater understanding of their hypothesised roles in eye function and development.

CHAPTER 5

Discussion

5.1 Evaluation of the Digital Differential Display approach

A Digital Differential Display (DDD) approach was taken to identify novel differentially expressed genes and was successful for this purpose, with the discovery of a number of novel genes that have been shown to be preferentially expressed by RT-PCR, and in some cases by *in situ* hybridisation at embryonic stages.

RT-PCR showed eye specific expression for four of the 28 sequences that were successfully amplified and a further 11 were confirmed to be preferentially expressed, with representation in the eye and three or less other tissues. Only one sequence was not expressed in the eye at any tested stage, showing that the presence of ESTs is a reasonable indicator of expression detectable by RT-PCR. Twelve sequences are expressed more widely, therefore the occurrence of false positives for differential expression is less than half. The DDD success rate lies within the range seen using subtractive techniques to elucidate novel genes preferentially expressed in the eye and is generally better than that seen in large scale *in situ* hybridisation analysis and in other *in silico* analysis. Previously, DDD has been used to find differentially expressed genes with respect to different cancer types and comparing cancerous to normal tissue, with only a tenth of tested novel genes showing specificity, compared to four of 15 (27%) in this study (Scheurle et al., 2000). If sequences that were later found to represent known genes are removed from this comparison, two of 12 are still shown to be specifically expressed, therefore the success of DDD has enabled the discovery of tissue specific novel genes, as well as those that are preferentially expressed. However, this RT-PCR analysis has been carried out with a limited number of tissues, as with most confirmational techniques, and can not definitively prove tissue specificity.

For the possible reasons discussed earlier (sections 4.4 and 5.3), several *in situ* hybridisations failed to give a result. Despite this, the results presented here show eye expression at a minimum of one stage for nine of 11 novel sequences analysed using this approach. The process of DDD, applying conditions to novel clusters for further

investigation and subsequent RT-PCR confirmation was vindicated. Of the nine eye expressing clusters, six describe novel genes, while the remaining three have no ORF. However, it is possible that these are UTRs of novel genes.

The importance of a gene in tissue development or function where preferential expression is seen is based on the assumption that RNA levels are indicative of functional protein levels. Although this is irrelevant where genes are specifically expressed, preferential expression relies on this quantitative relationship between RNA and protein and this may not be an accurate evaluation for all genes. Nevertheless, RNA levels are reasonable indicators of novel protein levels and subsequent protein assays can be used to confirm this relationship.

5.2 Possible future investigation

As detailed earlier (chapter 3), the number of sequences in UniGene are increasing and this allows build construction to become more accurate. Future use of the DDD tool will therefore give more accurate results. Also, a more direct approach to finding novel genes that may be implicated in human eye development and function would be to use DDD results from the human UniGene build, although this would require the identification of mouse orthologues for subsequent in depth analysis. As well as a build that is likely to be more accurate, the process of analysing DDD output can be automated using a recent tool, known as Digital Extractor, which is freely available from the developers (Madden et al., 2003).

The information gained through the DDD approach and experimental analysis of expression has provided a basis for more investigation of the expressed sequences discovered. Further expression analysis of these sequences could provide more information about the temporal as well as the spatial aspect of expression,

particularly as RT-PCR expression was not always confirmed by *in situ* hybridisation. Additionally, an obvious next step would be downregulation or elimination of expression. This could either be carried out on the whole organism level, invoking a phenotype that may indicate gene function, or at the cellular level, using analysis of differential expression to assess the effect of downregulation on gene expression.

As well as general steps that can be taken with the set of sequences already identified, there are specific questions that can be asked of each. These are discussed here.

5.2.1 Mm.95741

Although this sequence displays eye expression, it has no ORF. However, it is part of a Genscan prediction and may therefore be an untranslated part of a gene. The general approaches of further *in situ* hybridisation could be applied to this sequence, particularly in the adult eye, as expression was seen at this stage by RT-PCR. However very little information about the nature of this gene is forthcoming, preventing a more detailed investigation.

5.2.2 *Tsd*

The Mm.159861 sequence could represent part of a transcript lying antisense to *Dst*, however it has no ORF. It may be interesting to investigate the impact of downregulating *Tsd* to assess the effect, if any, on *Dst* expression.

5.2.3 Mm.150838

The expression of Mm.150838 in the neonate retina is widespread, yet signal appears to be strong within certain individual cells. The horizontal cells may be one of these strongly expressing cell types. This possibility could be investigated by using colocalisation of Mm.150838 *in situ* signal with an antibody to a horizontal cell protein such as calbindin. Although there is no ORF, this may be an untranslated region of a gene. Further *in situ* hybridisation at different stages may provide more information about the nature of this differential expression within the retina.

5.2.4 *Trale*

Trale codes for a membrane protein. The interesting aspect of this gene is the signal seen with the control probe in a contrasting domain, with *Trale* expressing in the retina and antisense expression in the lens. *In situ* hybridisation between E11.5 and the neonate would be useful to further define these expression patterns temporally. As well as inferring function, a knockdown or knockout approach could be used to evaluate the impact on antisense expression under these conditions.

5.2.5 *Bret*

Further *in situ* hybridisation is required to assess the cellular nature of *Bret* expression. In particular, colocalisation experiments using the *Bret* probes and antibodies to proteins found in specific cells within the developing neural retina could be carried out to identify expressing cells. Similar analysis of adult brain, in which RT-PCR indicates there is expression, could also elucidate the cells that express *Bret*. Following this, a mouse knockout approach would be valuable to infer function from phenotype. As the BTB/POZ domain that this predicted protein contains has been implicated in transcriptional repression in other proteins, it may also be interesting to carry out differential expression analysis between normal and knockout in an effort to discover genes that may be repressed by the functional protein.

5.2.6 *Duf676*

Colocalisation of *Duf676* with a choroidal or RPE marker would be the next step in its analysis, to more precisely determine cell types exhibiting expression in existing *in situ* data. As RT-PCR results have indicated expression in the heart, if expression is within the the choroid, it would not be unreasonable to suggest this is blood-related expression that may not be connected to eye specific function. However RT-PCR data is preferential to the eye, brain, heart and testis and the expressing cells may be located in the RPE instead of the choroid, suggesting that this is not necessarily the case. Of added interest is the presence of a rare but conserved domain with unknown function. In order to gain clues about the function of this protein, a reasonable approach may be to create a knockout in mice or a morpholino knockdown of the gene in zebrafish.

5.2.7 *Tset*

Further *in situ* analysis would be required to analyse the expressing cells. *In situ* hybridisation on sections of early embryo and at later stages of development as well as in the adult brain could be performed to achieve this. The NAP domain predicted for this protein suggests a regulatory role in gene expression, providing the ideal opportunity for downregulation to gain insight into genes that may be affected by its function. Once expressing cells have been identified, the cell type in question could be cultured. Using RNAi specific to *Tset*, this gene could be downregulated and experimental and control cells could be subjected to microarray analysis to find genes whose expression correlates with the downregulation of *Tset*. It would be expected that genes lying downstream of *Tset* would be amongst this set of genes. This could be complemented by a mouse knockout approach to provide a phenotype. At a cellular level, the TSET protein could be localised to assess the possibility that it may function as a transcriptional regulator. TSET produced in *Escherichia coli* could be used to raise antibodies in rabbit or sheep, which should prove useful in assessing whether TSET is localised to the nucleus, consistent with its potential role as a repressor.

5.2.8 *Cln8l*

The expression of *Cln8l* ranging from E9.5 to adult suggests an important role in eye development and function, specifically in the RPE where signal was detected at E16.5 and in the adult, with signal in the developing retina at E10.5 being hypothesised as presumptive RPE expression. Expression is also seen in the brain, both embryonically by *in situ* analysis and in the adult by RT-PCR. Analysis of *Cln8l* could therefore be carried out in brain by *in situ* in an effort to identify expressing

cell types. RPE expression could be confirmed by colocalisation experiments with *Cln8l* and an antibody to an RPE specific protein such as rhodopsin. A mouse knockout or morpholino knockdown in zebrafish may be utilised to assess the impact on RPE development or function. It is possible that if an eye phenotype is observed, it may be comparable to the *mnd/mnd* mouse, which has a CLN8 truncation and displays retinal degeneration. As CLN8 is thought to play a role in lipid metabolism, CLN8L may have a similar function. If this is the case, it may be localised in the endoplasmic reticulum, therefore localisation of CLN8L in the cell could ascertain whether this is a possibility. CLN8L antibodies would be raised as described for TSET antibodies.

5.2.9 *Sapan*

The expression of *Sapan* implies that it represents a protein important in early nervous system development. Additionally, expression in the lens and the previous identification of a jellyfish crystallin as a saposin homologue suggest that this may be a novel crystallin. Further *in situ* hybridisations, particularly using adult lens, are required to assess lens expression at more stages than those already provided. Further analysis of otic vesicle and developing neural tube expression are also necessary to identify expressing cells in these developing regions. As the jellyfish crystallin is homologous to saposins, rather than prosaposin, it would suggest that SAPAN is cleaved to produce individual saposin-like proteins if it is indeed a novel crystallin. In order to assess this possibility, the lens-expressed protein could be isolated using antibodies to the saposin structures implied in SAPAN. This would be done as previously described for TSET. The antibody could localise SAPAN derived proteins. This would allow the cellular localisation of SAPAN to be determined, as well as permitting the isolation of the proteins in lens cells. Once isolated, crystallography of the products would elucidate the size of the proteins, indicating whether cleavage of SAPAN has occurred. A mouse knockout would serve to

provide clues to protein function by analysis of a possible phenotype. As *Sapan* is located in the opposite direction and in an intron of *Sorcs2*, it is possible that there is a co-regulation of expression between the two genes, although this would be more likely if they were antisense to each other. Differential expression analysis may be utilised to check the effect, if any, on *Sorcs2* expression, using RNAi to downregulate one gene and quantitative PCR to assess the effect on the other.

5.3 Conclusions

This thesis has shown that a relatively simple *in silico* approach followed by experimental expression analysis is effective at identifying interesting novel genes that are preferentially expressed in the eye. Resulting from this method are several genes that are worthy of further investigation and may be candidates for eye related diseases.

References

Adams,M.D., Kelley,J.M., Gocayne,J.D., Dubnick,M., Polymeropoulos,M.H., Xiao,H., Merril,C.R., Wu,A., Olde,B., Moreno,R.F., and . (1991). Complementary DNA sequencing: expressed sequence tags and human genome project. Science *252*, 1651-1656.

Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White,O., and . (1995). Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. Nature *377*, 3-174.

Agarwal,N., Swaroop,A., Zheng,K., Liou,J.D., O'Rourke,K., Graves,K.A., Gieser,L., Del Monte,M., and Yang-Feng,T.L. (1995). Expression and chromosomal localization of cDNA clones from an enriched human retinal pigment epithelial (RPE) cell line library: identification of two RPE-specific genes. Cytogenet. Cell Genet. *69*, 71-74.

Altschul,S.F., Gish,W., Miller,W., Myers,E.W., and Lipman,D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403-410.

Alwine,J.C., Kemp,D.J., and Stark,G.R. (1977). Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. Proc. Natl. Acad. Sci. U. S. A *74*, 5350-5354.

Anniss,A.M., Apostolopoulos,J., Dworkin,S., Purton,L.E., and Sparrow,R.L. (2002). An oxysterol-binding protein family identified in the mouse. DNA Cell Biol. *21*, 571-580.

Arnemann,J., Jakubiczka,S., Thuring,S., and Schmidtke,J. (1991). Cloning and sequence analysis of a human Y-chromosome-derived, testicular cDNA, TSPY. Genomics *11*, 108-114.

Ayyagari,R., Demirci,F.Y., Liu,J., Bingham,E.L., Stringham,H., Kakuk,L.E., Boehnke,M., Gorin,M.B., Richards,J.E., and Sieving,P.A. (2002). X-linked recessive atrophic macular degeneration from RPGR mutation. Genomics *80*, 166-171.

Azuma,N., Yamaguchi,Y., Handa,H., Hayakawa,M., Kanai,A., and Yamada,M. (1999). Missense mutation in the alternative splice region of the PAX6 gene in eye anomalies. Am. J. Hum. Genet. *65*, 656-663.

Baehr,W., Falk,J.D., Bugra,K., Triantafyllos,J.T., and McGinnis,J.F. (1988). Isolation and analysis of the mouse opsin gene. FEBS Lett. *238*, 253-256.

Bardwell,V.J. and Treisman,R. (1994). The POZ domain: a conserved protein-protein interaction motif. Genes Dev. *8*, 1664-1677.

Bedford,M., Arman,E., Orr-Urtreger,A., and Lonai,P. (1995). Analysis of the Hoxd-3 gene: structure and localization of its sense and natural antisense transcripts. DNA Cell Biol. *14*, 295-304.

Benson,D.A., Karsch-Mizrachi,I., Lipman,D.J., Ostell,J., and Wheeler,D.L. (2003). GenBank. Nucleic Acids Res. *31*, 23-27.

Bernier,G., Brown,A., Dalpe,G., De Repentigny,Y., Mathieu,M., and Kothary,R. (1995). Dystonin expression in the developing nervous system predominates in the neurons that degenerate in dystonia musculorum mutant mice. Mol. Cell Neurosci. *6*, 509-520.

Bernstein,S.L., Borst,D.E., Neuder,M.E., and Wong,P. (1996). Characterization of a human fovea cDNA library and regional differential gene expression in the human retina. Genomics *32*, 301-308.

Berry,V., Francis,P., Reddy,M.A., Collyer,D., Vithana,E., MacKay,I., Dawson,G., Carey,A.H., Moore,A., Bhattacharya,S.S., and Quinlan,R.A. (2001). Alpha-B crystallin gene (CRYAB) mutation causes dominant congenital posterior polar cataract in humans. Am. J. Hum. Genet. *69*, 1141-1145.

Berson,E.L. (1996). Retinitis pigmentosa: unfolding its mystery. Proc. Natl. Acad. Sci. U. S. A *93*, 4526-4528.

Bhat,S.P. and Nagineni,C.N. (1989). Alpha B subunit of lens-specific protein alpha-crystallin is present in other ocular and non-ocular tissues. Biochem. Biophys. Res. Commun. *158*, 319-325.

Bishop,J.O., Morton,J.G., Rosbash,M., and Richardson,M. (1974). Three abundance classes in HeLa cell messenger RNA. Nature *250*, 199-204.

Blackshaw,S., Fraioli,R.E., Furukawa,T., and Cepko,C.L. (2001). Comprehensive analysis of photoreceptor gene expression and the identification of candidate retinal disease genes. Cell *107*, 579-589.

Blin-Wakkach,C., Lezot,F., Ghoul-Mazgar,S., Hotton,D., Monteiro,S., Teillaud,C., Pibouin,L., Orestes-Cardoso,S., Papagerakis,P., Macdougall,M., Robert,B., and Berdal,A. (2001). Endogenous Msx1 antisense transcript: in vivo and in vitro evidences, structure, and potential involvement in skeleton development in mammals. Proc. Natl. Acad. Sci. U. S. A *98*, 7336-7341.

Bode,V.C. (1984). Ethylnitrosourea mutagenesis and the isolation of mutant alleles for specific genes located in the T region of mouse chromosome 17. Genetics *108*, 457-470.

Boguski,M.S., Lowe,T.M., and Tolstoshev,C.M. (1993). dbEST--database for "expressed sequence tags". Nat. Genet. *4*, 332-333.

Boguski,M.S. and Schuler,G.D. (1995). ESTablishing a human transcript map. Nat. Genet. *10*, 369-371.

Boguski,M.S., Tolstoshev,C.M., and Bassett,D.E., Jr. (1994). Gene discovery in dbEST. Science *265*, 1993-1994.

Bok,D. (1985). Retinal photoreceptor-pigment epithelium interactions. Friedenwald lecture. Invest Ophthalmol. Vis. Sci. *26*, 1659-1694.

Bok,D. (1993). The retinal pigment epithelium: a versatile partner in vision. J. Cell Sci. Suppl *17*, 189-195.

Bortoluzzi,S., d'Alessi,F., and Danieli,G.A. (2000). A novel resource for the study of genes expressed in the adult human retina. Invest Ophthalmol. Vis. Sci. *41*, 3305-3308.

Bouck,J., Yu,W., Gibbs,R., and Worley,K. (1999). Comparison of gene indexing databases. Trends Genet. *15*, 159-162.

Bovolenta,P., Mallamaci,A., Briata,P., Corte,G., and Boncinelli,E. (1997). Implication of OTX2 in pigment epithelium determination and neural retina differentiation. J. Neurosci. *17*, 4243-4252.

Boyle,D. and Takemoto,L. (1994). Characterization of the alpha-gamma and alpha-beta complex: evidence for an in vivo functional role of alpha-crystallin as a molecular chaperone. Exp. Eye Res. *58*, 9-15.

Brady,K.P., Rowe,L.B., Her,H., Stevens,T.J., Eppig,J., Sussman,D.J., Sikela,J., and Beier,D.R. (1997). Genetic mapping of 262 loci derived from expressed sequences in a murine interspecific cross using single-strand conformational polymorphism analysis. Genome Res. *7*, 1085-1093.

Brenner,S., Johnson,M., Bridgham,J., Golda,G., Lloyd,D.H., Johnson,D., Luo,S., McCurdy,S., Foy,M., Ewan,M., Roth,R., George,D., Eletr,S., Albrecht,G., Vermaas,E., Williams,S.R., Moon,K., Burcham,T., Pallas,M., DuBridge,R.B., Kirchner,J., Fearon,K., Mao,J., and Corcoran,K. (2000). Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nat. Biotechnol. *18*, 630-634.

Brett,D., Hanke,J., Lehmann,G., Haase,S., Delbruck,S., Krueger,S., Reich,J., and Bork,P. (2000). EST comparison indicates 38% of human mRNAs contain possible alternative splice forms. FEBS Lett. *474*, 83-86.

Brown,A., Bernier,G., Mathieu,M., Rossant,J., and Kothary,R. (1995). The mouse dystonia musculorum gene is a neural isoform of bullous pemphigoid antigen 1. Nat. Genet. *10*, 301-306.

Buraczynska,M., Mears,A.J., Zareparsi,S., Farjo,R., Filippova,E., Yuan,Y., MacNee,S.P., Hughes,B., and Swaroop,A. (2002). Gene expression profile of native human retinal pigment epithelium. Invest Ophthalmol. Vis. Sci. *43*, 603-607.

Cai,H., Howells,R.D., and Wagner,B.J. (1999). Identification of a novel gene product preferentially expressed in rat lens epithelial cells. Mol. Vis. *5*, 3.

Campana,W.M., Hiraiwa,M., Addison,K.C., and O'Brien,J.S. (1996). Induction of MAPK phosphorylation by prosaposin and prosaptide in PC12 cells. Biochem. Biophys. Res. Commun. *229*, 706-712.

Carninci,P., Waki,K., Shiraki,T., Konno,H., Shibata,K., Itoh,M., Aizawa,K., Arakawa,T., Ishii,Y., Sasaki,D., Bono,H., Kondo,S., Sugahara,Y., Saito,R., Osato,N., Fukuda,S., Sato,K., Watahiki,A., Hirozane-Kishikawa,T., Nakamura,M., Shibata,Y., Yasunishi,A., Kikuchi,N., Yoshiki,A., Kusakabe,M., Gustincich,S., Beisel,K., Pavan,W., Aidinis,V., Nakagawara,A., Held,W.A., Iwata,H., Kono,T., Nakauchi,H., Lyons,P., Wells,C., Hume,D.A., Fagiolini,M., Hensch,T.K., Brinkmeier,M., Camper,S., Hirota,J., Mombaerts,P., Muramatsu,M., Okazaki,Y., Kawai,J., and Hayashizaki,Y. (2003). Targeting a complex transcriptome: the construction of the mouse full-length cDNA encyclopedia. Genome Res. *13*, 1273-1289.

Carter,J.M., Hutcheson,A.M., and Quinlan,R.A. (1995). In vitro studies on the assembly properties of the lens proteins CP49, CP115: coassembly with alpha-crystallin but not with vimentin. Exp. Eye Res *60*, 181-192.

Carver,J.A., Nicholls,K.A., Aquilina,J.A., and Truscott,R.J. (1996). Age-related changes in bovine alpha-crystallin and high-molecular-weight protein. Exp. Eye Res. *63*, 639-647.

Chang,B., Smith,R.S., Hawes,N.L., Anderson,M.G., Zabaleta,A., Savinova,O., Roderick,T.H., Heckenlively,J.R., Davisson,M.T., and John,S.W. (1999). Interacting loci cause severe iris atrophy and glaucoma in DBA/2J mice. Nat. Genet. *21*, 405-409.

Chapple,J.P., Grayson,C., Hardcastle,A.J., Saliba,R.S., van der,S.J., and Cheetham,M.E. (2001). Unfolding retinal dystrophies: a role for molecular chaperones? Trends Mol. Med. *7*, 414-421.

Chauhan,B.K., Reed,N.A., Zhang,W., Duncan,M.K., Kilimann,M.W., and Cvekl,A. (2002). Identification of Genes Downstream of Pax6 in the Mouse Lens Using cDNA Microarrays. J. Biol. Chem. *277*, 11539-11548.

Choi,S., Hao,W., Chen,C.K., and Simon,M.I. (2001). Gene expression profiles of light-induced apoptosis in arrestin/rhodopsin kinase-deficient mouse retinas. Proc. Natl. Acad. Sci. U. S. A *98*, 13096-13101.

Coca-Prados,M., Escribano,J., and Ortego,J. (1999). Differential gene expression in the human ciliary epithelium. Prog. Retin. Eye Res. *18*, 403-429.

Cox,R.D. and Brown,S.D. (2003). Rodent models of genetic disease. Curr. Opin. Genet. Dev. *13*, 278-283.

Currie,M.G., Fok,K.F., Kato,J., Moore,R.J., Hamra,F.K., Duffin,K.L., and Smith,C.E. (1992). Guanylin: an endogenous activator of intestinal guanylate cyclase. Proc. Natl. Acad. Sci. U. S. A *89*, 947-951.

Danciger,M., Farber,D.B., Peyser,M., and Kozak,C.A. (1990). The gene for the beta-subunit of retinal transducin (Gnb-1) maps to distal mouse chromosome 4, and related sequences map to mouse chromosomes 5 and 8. Genomics *6*, 428-435.

Dasari,V.K., Goharderakhshan,R.Z., Perinchery,G., Li,L.C., Tanaka,Y., Alonzo,J., and Dahiya,R. (2001). Expression analysis of Y chromosome genes in human prostate cancer. J. Urol. *165*, 1335-1341.

Davidson,E.H. and Britten,R.J. (1979). Regulation of gene expression: possible role of repetitive sequences. Science *204*, 1052-1059.

de Jong,W.W., Hendriks,W., Mulders,J.W., and Bloemendal,H. (1989). Evolution of eye lens crystallins: the stress connection. Trends Biochem. Sci. *14*, 365-368.

den Hollander,A.I., van Driel,M.A., de Kok,Y.J., van de Pol,D.J., Hoyng,C.B., Brunner,H.G., Deutman,A.F., and Cremers,F.P. (1999). Isolation and mapping of novel candidate genes for retinal disorders using suppression subtractive hybridization. Genomics *58*, 240-249.

Deweindt,C., Albagli,O., Bernardin,F., Dhordain,P., Quief,S., Lantoine,D., Kerckaert,J.P., and Leprince,D. (1995). The LAZ3/BCL6 oncogene encodes a sequence-specific transcriptional inhibitor: a novel function for the BTB/POZ domain as an autonomous repressing domain. Cell Growth Differ. *6*, 1495-1503.

Diatchenko,L., Lau,Y.F., Campbell,A.P., Chenchik,A., Moqadam,F., Huang,B., Lukyanov,S., Lukyanov,K., Gurskaya,N., Sverdlov,E.D., and Siebert,P.D. (1996). Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. Proc. Natl. Acad. Sci. U. S A *93*, 6025-6030.

Diaz,E., Yang,Y.H., Ferreira,T., Loh,K.C., Okazaki,Y., Hayashizaki,Y., Tessier-Lavigne,M., Speed,T.P., and Ngai,J. (2003). Analysis of gene expression in the developing mouse retina. Proc. Natl. Acad. Sci. U. S. A.

Dowling,J., Yang,Y., Wollmann,R., Reichardt,L.F., and Fuchs,E. (1997). Developmental expression of BPAG1-n: insights into the spastic ataxia and gross neurologic degeneration in dystonia musculorum mice. Dev. Biol. *187*, 131-142.

Dubin,R.A., Wawrousek,E.F., and Piatigorsky,J. (1989). Expression of the murine alpha B-crystallin gene is not restricted to the lens. Mol. Cell Biol. *9*, 1083-1091.

Duchen,L.W., STRICH,S.J., and FALCONER,D.S. (1964). CLINICAL AND PATHOLOGICAL STUDIES OF AN HEREDITARY NEUROPATHY IN MICE (DYSTONIA MUSCULORUM). Brain *87*, 367-378.

Dyer,M.A., Livesey,F.J., Cepko,C.L., and Oliver,G. (2003). Prox1 function controls progenitor cell proliferation and horizontal cell genesis in the mammalian retina. Nat. Genet. *34*, 53-58.

Edman,C.F., Raymond,D.E., Wu,D.J., Tu,E., Sosnowski,R.G., Butler,W.F., Nerenberg,M., and Heller,M.J. (1997). Electric field directed nucleic acid hybridization on microchips. Nucleic Acids Res. *25*, 4907-4914.

Elliott,R.W., Sparkes,R.S., Mohandas,T., Grant,S.G., and McGinnis,J.F. (1990). Localization of the rhodopsin gene to the distal half of mouse chromosome 6. Genomics *6*, 635-644.

Escribano,J., Ortego,J., and Coca-Prados,M. (1995). Isolation and characterization of cell-specific cDNA clones from a subtractive library of the ocular ciliary body of a single normal human donor: transcription and synthesis of plasma proteins. J. Biochem. (Tokyo) *118*, 921-931.

Fantes,J.A., Bickmore,W.A., Fletcher,J.M., Ballesta,F., Hanson,I.M., and van,H., V (1992). Submicroscopic deletions at the WAGR locus, revealed by nonradioactive in situ hybridization. Am. J. Hum. Genet. *51*, 1286-1294.

Feldkaemper,M.P., Wang,H.Y., and Schaeffel,F. (2000). Changes in retinal and choroidal gene expression during development of refractive errors in chicks. Invest Ophthalmol. Vis. Sci. *41*, 1623-1628.

Freeman,W.M., Walker,S.J., and Vrana,K.E. (1999). Quantitative RT-PCR: pitfalls and potential. Biotechniques *26*, 112-115.

Freeman,W.M., Robertson,D.J., and Vrana,K.E. (2000). Fundamentals of DNA hybridization arrays for gene expression analysis. Biotechniques *29*, 1042-1055.

Friedman,J., Trahey,M., and Weissman,I. (1993). Cloning and characterization of cyclophilin C-associated protein: a candidate natural cellular ligand for cyclophilin C. Proc. Natl. Acad. Sci. U. S. A *90*, 6815-6819.

Fryer,R.M., Randall,J., Yoshida,T., Hsiao,L.L., Blumenstock,J., Jensen,K.E., Dimofte,T., Jensen,R.V., and Gullans,S.R. (2002). Global analysis of gene expression: methods, interpretation, and pitfalls. Exp. Nephrol. *10*, 64-74.

Fujita,N., Suzuki,K., Vanier,M.T., Popko,B., Maeda,N., Klein,A., Henseler,M., Sandhoff,K., Nakayasu,H., and Suzuki,K. (1996). Targeted disruption of the mouse sphingolipid activator protein gene: a complex phenotype, including severe leukodystrophy and wide-spread storage of multiple sphingolipids. Hum. Mol. Genet. *5*, 711-725.

Garriga,P. and Manyosa,J. (2002). The eye photoreceptor protein rhodopsin. Structural implications for retinal disease. FEBS Lett. *528*, 17-22.

Gawantka,V., Pollet,N., Delius,H., Vingron,M., Pfister,R., Nitsch,R., Blumenstock,C., and Niehrs,C. (1998). Gene expression screening in Xenopus identifies molecular pathways, predicts gene function and provides a global view of embryonic patterning. Mech. Dev. *77*, 95-141.

Gibbs,R.A. (1997). Hares and tortoises in the race to sequence the human genome: expectations and realities. Trends Genet. *13*, 381-383.

Glaser,T., Jepeal,L., Edwards,J.G., Young,S.R., Favor,J., and Maas,R.L. (1994). PAX6 gene dosage effect in a family with congenital cataracts, aniridia, anophthalmia and central nervous system defects. Nat. Genet. *7*, 463-471.

Gonzalez,P., Epstein,D.L., and Borras,T. (2000a). Characterization of gene expression in human trabecular meshwork using single-pass sequencing of 1060 clones. Invest Ophthalmol. Vis. Sci. *41*, 3678-3693.

Gonzalez,P., Epstein,D.L., and Borras,T. (2000b). Genes upregulated in the human trabecular meshwork in response to elevated intraocular pressure. Invest Ophthalmol. Vis. Sci. *41*, 352-361.

Goring,D.R., Breitman,M.L., and Tsui,L.C. (1992). Temporal regulation of six crystallin transcripts during mouse lens development. Exp. Eye Res. *54*, 785-795.

Gottsch,J.D., Bowers,A.L., Margulies,E.H., Seitzman,G.D., Kim,S.W., Saha,S., Jun,A.S., Stark,W.J., and Liu,S.H. (2003). Serial analysis of gene expression in the corneal endothelium of Fuchs' dystrophy. Invest Ophthalmol. Vis. Sci. *44*, 594-599.

Granadino,B., Gallardo,M.E., Lopez-Rios,J., Sanz,R., Ramos,C., Ayuso,C., Bovolenta,P., and Rodriguez,d.C. (1999). Genomic cloning, structure, expression pattern, and chromosomal location of the human SIX3 gene. Genomics *55*, 100-105.

Graw,J., Klopp,N., Neuhauser-Klaus,A., Favor,J., and Loster,J. (2002). Crygf(Rop): the first mutation in the Crygf gene causing a unique radial lens opacity. Invest Ophthalmol. Vis. Sci. *43*, 2998-3002.

Gridley,T., Soriano,P., and Jaenisch,R. (1987). Insertional mutagenesis in mice. Trends in Genetics *3*, 162-166.

Grindley,J.C., Davidson,D.R., and Hill,R.E. (1995). The role of Pax-6 in eye and nasal development. Development *121*, 1433-1442.

Hackam,A.S., Bradford,R.L., Bakhru,R.N., Shah,R.M., Farkas,R., Zack,D.J., and Adler,R. (2003). Gene discovery in the embryonic chick retina. Mol. Vis. *9*, 262-276.

Haeseleer,F., Jang,G.F., Imanishi,Y., Driessen,C.A., Matsumura,M., Nelson,P.S., and Palczewski,K. (2002). Dual-substrate specificity short chain retinol dehydrogenases from the vertebrate retina. J. Biol. Chem. *277*, 45537-45546.

Halder,G., Callaerts,P., and Gehring,W.J. (1995). Induction of ectopic eyes by targeted expression of the eyeless gene in Drosophila. Science *267*, 1788-1792.

Hedrick,S.M., Cohen,D.I., Nielsen,E.A., and Davis,M.M. (1984). Isolation of cDNA clones encoding T cell-specific membrane-associated proteins. Nature *308*, 149-153.

Hernando,N., Martin-Alonso,J.M., Ghosh,S., and Coca-Prados,M. (1992). Isolation of a cDNA encoding a glutathione S-transferase (GST) class-pi from the bovine ocular ciliary epithelium. Exp. Eye Res. *55*, 711-718.

Herrmann,B.G. (1991). Expression pattern of the Brachyury gene in whole-mount TWis/TWis mutant embryos. Development *113*, 913-917.

Hess,J.F., Casselman,J.T., Kong,A.P., and FitzGerald,P.G. (1998). Primary sequence, secondary structure, gene structure, and assembly properties suggests that the lens-specific cytoskeletal protein filensin represents a novel class of intermediate filament protein. Exp. Eye Res. *66*, 625-644.

Hill,R.E., Favor,J., Hogan,B.L., Ton,C.C., Saunders,G.F., Hanson,I.M., Prosser,J., Jordan,T., Hastie,N.D., and van,H., V (1991). Mouse small eye results from mutations in a paired-like homeobox-containing gene. Nature *354*, 522-525.

Hillier,L.D., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B., Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T., Schellenberg,K., Marra,M., and . (1996). Generation and analysis of 280,000 human expressed sequence tags. Genome Res. *6*, 807-828.

Holtschmidt,H., Sandhoff,K., Kwon,H.Y., Harzer,K., Nakano,T., and Suzuki,K. (1991). Sulfatide activator protein. Alternative splicing that generates three mRNAs and a newly found mutation responsible for a clinical disease. J. Biol. Chem. *266*, 7556-7560.

Hong,D.H., Pawlyk,B.S., Shang,J., Sandberg,M.A., Berson,E.L., and Li,T. (2000). A retinitis pigmentosa GTPase regulator (RPGR)-deficient mouse model for X-linked retinitis pigmentosa (RP3). Proc. Natl. Acad. Sci. U. S. A *97*, 3649-3654.

Houlgatte,R., Mariage-Samson,R., Duprat,S., Tessier,A., Bentolila,S., Lamy,B., and Auffray,C. (1995). The Genexpress Index: a resource for gene discovery and the genic map of the human genome. Genome Res. *5*, 272-304.

Howes,K., Bronson,J.D., Dang,Y.L., Li,N., Zhang,K., Ruiz,C., Helekar,B., Lee,M., Subbaraya,I., Kolb,H., Chen,J., and Baehr,W. (1998). Gene array and expression of mouse retina guanylate cyclase activating proteins 1 and 2. Invest Ophthalmol. Vis. Sci. *39*, 867-875.

Hulkova,H., Cervenkova,M., Ledvinova,J., Tochackova,M., Hrebicek,M., Poupetova,H., Befekadu,A., Berna,L., Paton,B.C., Harzer,K., Boor,A., Smid,F., and Elleder,M. (2001). A novel mutation in the coding region of the prosaposin gene leads to a complete deficiency of prosaposin and saposins, and is associated with a complex sphingolipidosis dominated by lactosylceramide accumulation. Hum. Mol. Genet. *10*, 927-940.

Humphries,M.M., Rancourt,D., Farrar,G.J., Kenna,P., Hazel,M., Bush,R.A., Sieving,P.A., Sheils,D.M., McNally,N., Creighton,P., Erven,A., Boros,A., Gulya,K.,

Capecchi,M.R., and Humphries,P. (1997). Retinopathy induced in mice by targeted disruption of the rhodopsin gene. Nat. Genet. *15*, 216-219.

Huynh,K.D. and Bardwell,V.J. (1998). The BCL-6 POZ domain and other POZ domains interact with the co-repressors N-CoR and SMRT. Oncogene *17*, 2473-2484.

Iacobelli,S., Arno,E., D'Orazio,A., and Coletti,G. (1986). Detection of antigens recognized by a novel monoclonal antibody in tissue and serum from patients with breast cancer. Cancer Res. *46*, 3005-3010.

Iacobelli,S., Arno,E., Sismondi,P., Natoli,C., Gentiloni,N., Scambia,G., Giai,M., Cortese,P., Panici,P.B., and Mancuso,S. (1988). Measurement of a breast cancer associated antigen detected by monoclonal antibody SP-2 in sera of cancer patients. Breast Cancer Res. Treat. *11*, 19-30.

Iacobelli,S., Bucci,I., D'Egidio,M., Giuliani,C., Natoli,C., Tinari,N., Rubistein,M., and Schlessinger,J. (1993). Purification and characterization of a 90 kDa protein released from human tumors and tumor cell lines. FEBS Lett. *319*, 59-65.

Jacobson,S.G., Cideciyan,A.V., Kemp,C.M., Sheffield,V.C., and Stone,E.M. (1996). Photoreceptor function in heterozygotes with insertion or deletion mutations in the RDS gene. Invest Ophthalmol. Vis. Sci. *37*, 1662-1674.

Jaenisch,R. (1988). Transgenic animals. Science *240*, 1468-1474.

Jin,S., Cornwall,M.C., and Oprian,D.D. (2003). Opsin activation as a cause of congenital night blindness. Nat. Neurosci. *6*, 731-735.

Jomary,C., Neal,M.J., and Jones,S.E. (2001). Characterization of cell death pathways in murine retinal neurodegeneration implicates cytochrome c release, caspase activation, and bid cleavage. Mol. Cell Neurosci. *18*, 335-346.

Jones,P.G., Lombardi,S.J., and Cockett,M.I. (1998). Cloning and tissue distribution of the human G protein beta 5 cDNA. Biochim. Biophys. Acta *1402*, 288-291.

Jones,S.E., Jomary,C., Grist,J., Stewart,H.J., and Neal,M.J. (2000a). Altered expression of secreted frizzled-related protein-2 in retinitis pigmentosa retinas. Invest Ophthalmol. Vis. Sci. *41*, 1297-1301.

Jones,S.E., Jomary,C., Grist,J., Stewart,H.J., and Neal,M.J. (2000b). Identification by array screening of altered nm23-M2/PuF mRNA expression in mouse retinal degeneration. Mol. Cell Biol. Res. Commun. *4*, 20-25.

Joussen,A.M., Huang,S., Poulaki,V., Camphausen,K., Beecken,W.D., Kirchhof,B., and Adamis,A.P. (2001). In vivo retinal gene expression in early diabetes. Invest Ophthalmol. Vis. Sci. *42*, 3047-3057.

Jun,A.S., Liu,S.H., Koo,E.H., Do,D.V., Stark,W.J., and Gottsch,J.D. (2001). Microarray analysis of gene expression in human donor corneas. Arch. Ophthalmol. *119*, 1629-1634.

Jursky,F. and Nelson,N. (1999). Developmental expression of the neurotransmitter transporter GAT3. J. Neurosci. Res. *55*, 394-399.

Kafatos,F.C., Jones,C.W., and Efstratiadis,A. (1979). Determination of nucleic acid sequence homologies and relative concentrations by a dot hybridization procedure. Nucleic Acids Res. *7*, 1541-1552.

Kajiwara,K., Sandberg,M.A., Berson,E.L., and Dryja,T.P. (1993). A null mutation in the human peripherin/RDS gene in a family with autosomal dominant retinitis punctata albescens. Nat. Genet. *3*, 208-212.

Kantorow,M., Kays,T., Horwitz,J., Huang,Q., Sun,J., Piatigorsky,J., and Carper,D. (1998). Differential display detects altered gene expression between cataractous and normal human lenses. Invest Ophthalmol. Vis. Sci. *39*, 2344-2354.

Katsanis,N., Worley,K.C., Gonzalez,G., Ansley,S.J., and Lupski,J.R. (2002). A computational/functional genomics approach for the enrichment of the retinal transcriptome and the identification of positional candidate retinopathy genes. Proc. Natl. Acad. Sci. U. S. A *99*, 14326-14331.

Kerrison,J.B., Arnould,V.J., Barmada,M.M., Koenekoop,R.K., Schmeckpeper,B.J., and Maumenee,I.H. (1996). A gene for autosomal dominant congenital nystagmus localizes to 6p12. Genomics *33*, 523-526.

Kimelman,D. and Kirschner,M.W. (1989). An antisense mRNA directs the covalent modification of the transcript encoding fibroblast growth factor in Xenopus oocytes. Cell *59*, 687-696.

Kirschner,R., Rosenberg,T., Schultz-Heienbrok,R., Lenzner,S., Feil,S., Roepman,R., Cremers,F.P., Ropers,H.H., and Berger,W. (1999). RPGR transcription studies in mouse and human tissues reveal a retina-specific isoform that is disrupted in a patient with X-linked retinitis pigmentosa. Hum. Mol. Genet. *8*, 1571-1578.

Kiyosawa,H., Yamanaka,I., Osato,N., Kondo,S., and Hayashizaki,Y. (2003). Antisense transcripts with FANTOM2 clone set and their implications for gene regulation. Genome Res. *13*, 1324-1334.

Knee,R.S., Pitcher,S.E., and Murphy,P.R. (1994). Basic fibroblast growth factor sense (FGF) and antisense (gfg) RNA transcripts are expressed in unfertilized human oocytes and in differentiated adult tissues. Biochem. Biophys. Res. Commun. *205*, 577-583.

Ko,M.S., Kitchen,J.R., Wang,X., Threat,T.A., Wang,X., Hasegawa,A., Sun,T., Grahovac,M.J., Kargul,G.J., Lim,M.K., Cui,Y., Sano,Y., Tanaka,T., Liang,Y., Mason,S., Paonessa,P.D., Sauls,A.D., DePalma,G.E., Sharara,R., Rowe,L.B., Eppig,J., Morrell,C., and Doi,H. (2000). Large-scale cDNA analysis reveals phased

gene expression patterns during preimplantation mouse development. Development *127*, 1737-1749.

Koths,K., Taylor,E., Halenbeck,R., Casipit,C., and Wang,A. (1993). Cloning and characterization of a human Mac-2-binding protein, a new member of the superfamily defined by the macrophage scavenger receptor cysteine-rich domain. J. Biol. Chem. *268*, 14245-14249.

Kumar,S., Tamura,K., Jakobsen,I.B., and Nei,M. (2001). MEGA2: molecular evolutionary genetics analysis software. Bioinformatics. *17*, 1244-1245.

Lagutin,O., Zhu,C.C., Furuta,Y., Rowitch,D.H., McMahon,A.P., and Oliver,G. (2001). Six3 promotes the formation of ectopic optic vesicle-like structures in mouse embryos. Dev. Dyn. *221*, 342-349.

Lamb,T.D. (1996). Gain and kinetics of activation in the G-protein cascade of phototransduction. Proc. Natl. Acad. Sci. U. S A *93*, 566-570.

Lander,E.S., Linton,L.M., Birren,B., Nusbaum,C., Zody,M.C., Baldwin,J., Devon,K., Dewar,K., Doyle,M., FitzHugh,W., Funke,R., Gage,D., Harris,K., Heaford,A., Howland,J., Kann,L., Lehoczky,J., Levine,R., McEwan,P., McKernan,K., Meldrim,J., Mesirov,J.P., Miranda,C., Morris,W., Naylor,J., Raymond,C., Rosetti,M., Santos,R., Sheridan,A., Sougnez,C., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Wyman,D., Rogers,J., Sulston,J., Ainscough,R., Beck,S., Bentley,D., Burton,J., Clee,C., Carter,N., Coulson,A., Deadman,R., Deloukas,P., Dunham,A., Dunham,I., Durbin,R., French,L., Grafham,D., Gregory,S., Hubbard,T., Humphray,S., Hunt,A., Jones,M., Lloyd,C., McMurray,A., Matthews,L., Mercer,S., Milne,S., Mullikin,J.C., Mungall,A., Plumb,R., Ross,M., Shownkeen,R., Sims,S., Waterston,R.H., Wilson,R.K., Hillier,L.W., McPherson,J.D., Marra,M.A., Mardis,E.R., Fulton,L.A., Chinwalla,A.T., Pepin,K.H., Gish,W.R., Chissoe,S.L., Wendl,M.C., Delehaunty,K.D., Miner,T.L., Delehaunty,A., Kramer,J.B., Cook,L.L., Fulton,R.S., Johnson,D.L., Minx,P.J., Clifton,S.W., Hawkins,T., Branscomb,E., Predki,P., Richardson,P., Wenning,S., Slezak,T., Doggett,N., Cheng,J.F., Olsen,A., Lucas,S., Elkin,C., Uberbacher,E., Frazier,M., Gibbs,R.A., Muzny,D.M., Scherer,S.E., Bouck,J.B., Sodergren,E.J., Worley,K.C., Rives,C.M., Gorrell,J.H., Metzker,M.L., Naylor,S.L., Kucherlapati,R.S., Nelson,D.L., Weinstock,G.M., Sakaki,Y., Fujiyama,A., Hattori,M., Yada,T., Toyoda,A., Itoh,T., Kawagoe,C., Watanabe,H., Totoki,Y., Taylor,T., Weissenbach,J., Heilig,R., Saurin,W., Artiguenave,F., Brottier,P., Bruls,T., Pelletier,E., Robert,C., Wincker,P., Smith,D.R., Doucette-Stamm,L., Rubenfield,M., Weinstock,K., Lee,H.M., Dubois,J., Rosenthal,A., Platzer,M., Nyakatura,G., Taudien,S., Rump,A., Yang,H., Yu,J., Wang,J., Huang,G., Gu,J., Hood,L., Rowen,L., Madan,A., Qin,S., Davis,R.W., Federspiel,N.A., Abola,A.P., Proctor,M.J., Myers,R.M., Schmutz,J., Dickson,M., Grimwood,J., Cox,D.R., Olson,M.V., Kaul,R., Raymond,C., Shimizu,N., Kawasaki,K., Minoshima,S., Evans,G.A., Athanasiou,M., Schultz,R., Roe,B.A., Chen,F., Pan,H., Ramser,J., Lehrach,H., Reinhardt,R., McCombie,W.R., de la,B.M., Dedhia,N., Blocker,H., Hornischer,K., Nordsiek,G., Agarwala,R., Aravind,L., Bailey,J.A., Bateman,A., Batzoglou,S., Birney,E., Bork,P., Brown,D.G., Burge,C.B., Cerutti,L., Chen,H.C., Church,D., Clamp,M., Copley,R.R., Doerks,T., Eddy,S.R.,

Eichler,E.E., Furey,T.S., Galagan,J., Gilbert,J.G., Harmon,C., Hayashizaki,Y., Haussler,D., Hermjakob,H., Hokamp,K., Jang,W., Johnson,L.S., Jones,T.A., Kasif,S., Kaspryzk,A., Kennedy,S., Kent,W.J., Kitts,P., Koonin,E.V., Korf,I., Kulp,D., Lancet,D., Lowe,T.M., McLysaght,A., Mikkelsen,T., Moran,J.V., Mulder,N., Pollara,V.J., Ponting,C.P., Schuler,G., Schultz,J., Slater,G., Smit,A.F., Stupka,E., Szustakowski,J., Thierry-Mieg,D., Thierry-Mieg,J., Wagner,L., Wallis,J., Wheeler,R., Williams,A., Wolf,Y.I., Wolfe,K.H., Yang,S.P., Yeh,R.F., Collins,F., Guyer,M.S., Peterson,J., Felsenfeld,A., Wetterstrand,K.A., Patrinos,A., Morgan,M.J., Szustakowki,J., de Jong,P., Catanese,J.J., Osoegawa,K., Shizuya,H., and Choi,S. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Lau,Y.F. and Zhang,J. (2000). Expression analysis of thirty one Y chromosome genes in human prostate cancer. Mol. Carcinog. *27*, 308-321.

Lebioda,L. and Stec,B. (1991). Mapping of isozymic differences in enolase. Int. J. Biol. Macromol. *13*, 97-100.

Lefler,W.H., Wadsworth,J.A., and Sidbury,J.B., Jr. (1971). Hereditary macular degeneration and amino-aciduria. Am. J. Ophthalmol. *1*, 224-230.

Lem,J., Krasnoperova,N.V., Calvert,P.D., Kosaras,B., Cameron,D.A., Nicolo,M., Makino,C.L., and Sidman,R.L. (1999). Morphological, physiological, and biochemical changes in rhodopsin knockout mice. Proc. Natl. Acad. Sci. U. S. A *96*, 736-741.

Lengler,J. and Graw,J. (2001). Regulation of the human SIX3 gene promoter. Biochem. Biophys. Res. Commun. *287*, 372-376.

Lennon,G., Auffray,C., Polymeropoulos,M., and Soares,M.B. (1996). The I.M.A.G.E. Consortium: an integrated molecular analysis of genomes and their expression. Genomics *33*, 151-152.

Levin,L.A. and Geszvain,K.M. (1998). Expression of ceruloplasmin in the retina: induction after optic nerve crush. Invest Ophthalmol. Vis. Sci. *39*, 157-163.

Levine,M.A., Smallwood,P.M., Moen,P.T., Jr., Helman,L.J., and Ahn,T.G. (1990). Molecular cloning of beta 3 subunit, a third form of the G protein beta-subunit polypeptide. Proc. Natl. Acad. Sci. U. S. A *87*, 2329-2333.

Li,A.W. and Murphy,P.R. (2000). Expression of alternatively spliced FGF-2 antisense RNA transcripts in the central nervous system: regulation of FGF-2 mRNA translation. Mol. Cell Endocrinol. *170*, 233-242.

Liang,P. and Pardee,A.B. (1992). Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science *257*, 967-971.

Lin,C.T. and Sargan,D.R. (2001). Generation and analysis of canine retinal ESTs: isolation and expression of retina-specific gene transcripts. Biochem. Biophys. Res Commun. *282*, 394-403.

Lipshutz,R.J., Fodor,S.P., Gingeras,T.R., and Lockhart,D.J. (1999). High density synthetic oligonucleotide arrays. Nat. Genet. *21*, 20-24.

Litt,M., Kramer,P., LaMorticella,D.M., Murphey,W., Lovrien,E.W., and Weleber,R.G. (1998). Autosomal dominant congenital cataract associated with a missense mutation in the human alpha crystallin gene CRYAA. Hum. Mol. Genet. *7*, 471-474.

Liu,Q.R., Lopez-Corcuera,B., Mandiyan,S., Nelson,H., and Nelson,N. (1993). Molecular characterization of four pharmacologically distinct gamma-aminobutyric acid transporters in mouse brain [corrected]. J. Biol. Chem. *268*, 2106-2112.

Livesey,F.J., Furukawa,T., Steffen,M.A., Church,G.M., and Cepko,C.L. (2000). Microarray analysis of the transcriptional network controlled by the photoreceptor homeobox gene Crx. Curr. Biol. *10*, 301-310.

Lu,C., Kasik,J., Stephan,D.A., Yang,S., Sperling,M.A., and Menon,R.K. (2001). Grtp1, a novel gene regulated by growth hormone. Endocrinology *142*, 4568-4571.

Ma,Q. and Whitlock,J.P., Jr. (1997). A novel cytoplasmic protein that interacts with the Ah receptor, contains tetratricopeptide repeat motifs, and augments the transcriptional response to 2,3,7,8-tetrachlorodibenzo-p-dioxin. J. Biol. Chem. *272*, 8878-8884.

Mackay,D.S., Andley,U.P., and Shiels,A. (2003). Cell death triggered by a novel mutation in the alphaA-crystallin gene underlies autosomal dominant cataract linked to chromosome 21q. Eur. J. Hum. Genet. *11*, 784-793.

Madden,S.F., O'Donovan,B., Furney,S.J., Brady,H.R., Silvestre,G., and Doran,P.P. (2003). Digital extractor: analysis of digital differential display output. Bioinformatics. *19*, 1594-1595.

Malone,K., Sohocki,M.M., Sullivan,L.S., and Daiger,S.P. (1999). Identifying and mapping novel retinal-expressed ESTs from humans. Mol. Vis. *5*, 5.

Marra,M., Hillier,L., Kucaba,T., Allen,M., Barstead,R., Beck,C., Blistain,A., Bonaldo,M., Bowers,Y., Bowles,L., Cardenas,M., Chamberlain,A., Chappell,J., Clifton,S., Favello,A., Geisel,S., Gibbons,M., Harvey,N., Hill,F., Jackson,Y., Kohn,S., Lennon,G., Mardis,E., Martin,J., Waterston,R., and . (1999). An encyclopedia of mouse genes. Nat. Genet. *21*, 191-194.

Matsubara,K. and Okubo,K. (1993). cDNA analyses in the human genome project. Gene *135*, 265-274.

Matsuda,J., Vanier,M.T., Saito,Y., Tohyama,J., Suzuki,K., and Suzuki,K. (2001). A mutation in the saposin A domain of the sphingolipid activator protein (prosaposin) gene results in a late-onset, chronic form of globoid cell leukodystrophy in the mouse. Hum. Mol. Genet. *10*, 1191-1199.

Matsuo,I., Kuratani,S., Kimura,C., Takeda,N., and Aizawa,S. (1995). Mouse Otx2 functions in the formation and patterning of rostral head. Genes Dev. *9*, 2646-2658.

Mavlyutov,T.A., Zhao,H., and Ferreira,P.A. (2002). Species-specific subcellular localization of RPGR and RPGRIP isoforms: implications for the phenotypic variability of congenital retinopathies among species. Hum. Mol. Genet. *11*, 1899-1907.

Mayor,C., Brudno,M., Schwartz,J.R., Poliakov,A., Rubin,E.M., Frazer,K.A., Pachter,L.S., and Dubchak,I. (2000). VISTA : visualizing global DNA sequence alignments of arbitrary length. Bioinformatics. *16*, 1046-1047.

McAvoy,J.W., Chamberlain,C.G., de Iongh,R.U., Hales,A.M., and Lovicu,F.J. (1999). Lens development. Eye *13 (Pt 3b)*, 425-437.

Megy,K., Audic,S., and Claverie,J.M. (2003). Positional clustering of differentially expressed genes on human chromosomes 20, 21 and 22. Genome Biol. *4*, 1.

Meindl,A., Dry,K., Herrmann,K., Manson,F., Ciccodicola,A., Edgar,A., Carvalho,M.R., Achatz,H., Hellebrand,H., Lennon,A., Migliaccio,C., Porter,K., Zrenner,E., Bird,A., Jay,M., Lorenz,B., Wittwer,B., D'Urso,M., Meitinger,T., and Wright,A. (1996). A gene (RPGR) with homology to the RCC1 guanine nucleotide exchange factor is mutated in X-linked retinitis pigmentosa (RP3). Nat. Genet. *13*, 35-42.

Mendez,A., Burns,M.E., Sokal,I., Dizhoor,A.M., Baehr,W., Palczewski,K., Baylor,D.A., and Chen,J. (2001). Role of guanylate cyclase-activating proteins (GCAPs) in setting the flash sensitivity of rod photoreceptors. Proc. Natl. Acad. Sci. U. S. A *98*, 9948-9953.

Messer,A. and Flaherty,L. (1986). Autosomal dominance in a late-onset motor neuron disease in the mouse. J. Neurogenet. *3*, 345-355.

Minoshima,S., Amagai,M., Kudoh,J., Fukuyama,R., Hashimoto,T., Nishikawa,T., and Shimizu,N. (1991). Localization of the human gene for 230-kDal bullous pemphigoid autoantigen (BPAG1) to chromosome 6pter----q15. Cytogenet. Cell Genet. *57*, 30-32.

Mitsui,K., Tokuzawa,Y., Itoh,H., Segawa,K., Murakami,M., Takahashi,K., Maruyama,M., Maeda,M., and Yamanaka,S. (2003). The Homeoprotein Nanog Is Required for Maintenance of Pluripotency in Mouse Epiblast and ES Cells. Cell *113*, 631-642.

Moreira,E.F., Jaworski,C., Li,A., and Rodriguez,I.R. (2001). Molecular and biochemical characterization of a novel oxysterol-binding protein (OSBP2) highly expressed in retina. J. Biol. Chem. *276*, 18570-18578.

Muchowski,P.J., Valdez,M.M., and Clark,J.I. (1999). AlphaB-crystallin selectively targets intermediate filament proteins during thermal stress. Invest Ophthalmol. Vis. Sci. *40*, 951-958.

Mullis,K., Faloona,F., Scharf,S., Saiki,R., Horn,G., and Erlich,H. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harb. Symp. Quant. Biol. *51 Pt 1*, 263-273.

Murphy,P.R. and Knee,R.S. (1994). Identification and characterization of an antisense RNA transcript (gfg) from the human basic fibroblast growth factor gene. Mol. Endocrinol. *8*, 852-859.

Nathans,J., Thomas,D., and Hogness,D.S. (1986). Molecular genetics of human color vision: the genes encoding blue, green, and red pigments. Science *232*, 193-202.

Neidhardt,L., Gasca,S., Wertz,K., Obermayr,F., Worpenberg,S., Lehrach,H., and Herrmann,B.G. (2000). Large-scale screen for genes controlling mammalian embryogenesis, using high-throughput gene expression analysis in mouse embryos. Mech. Dev. *98*, 77-94.

Nguyen,C., Rocha,D., Granjeaud,S., Baldit,M., Bernard,K., Naquet,P., and Jordan,B.R. (1995). Differential gene expression in the murine thymus assayed by quantitative hybridization of arrayed cDNA clones. Genomics *29*, 207-216.

Nielsen,K., Birkenkamp-Demtroder,K., Ehlers,N., and Orntoft,T.F. (2003). Identification of differentially expressed genes in keratoconus epithelium analyzed on microarrays. Invest Ophthalmol. Vis. Sci. *44*, 2466-2476.

Nishida,K., Adachi,W., Shimizu-Matsumoto,A., Kinoshita,S., Mizuno,K., Matsubara,K., and Okubo,K. (1996). A gene expression profile of human corneal epithelium and the isolation of human keratin 12 cDNA. Invest Ophthalmol. Vis. Sci. *37*, 1800-1809.

O'Brien,J.S., Carson,G.S., Seo,H.C., Hiraiwa,M., and Kishimoto,Y. (1994). Identification of prosaposin as a neurotrophic factor. Proc. Natl. Acad. Sci. U. S. A *91*, 9593-9596.

O'Brien,J.S., Carson,G.S., Seo,H.C., Hiraiwa,M., Weiler,S., Tomich,J.M., Barranger,J.A., Kahn,M., Azuma,N., and Kishimoto,Y. (1995). Identification of the neurotrophic factor sequence of prosaposin. FASEB J. *9*, 681-685.

O'Brien,J.S. and Kishimoto,Y. (1991). Saposin proteins: structure, function, and role in human lysosomal storage disorders. FASEB J. *5*, 301-308.

Ohki,T., Hongo,S., Nakada,N., Maeda,A., and Takeda,M. (2002). Inhibition of neurite outgrowth by reduced level of NDRG4 protein in antisense transfected PC12 cells. Brain Res. Dev. Brain Res. *135*, 55-63.

Okubo,K., Hori,N., Matoba,R., Niiyama,T., Fukushima,A., Kojima,Y., and Matsubara,K. (1992). Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression. Nat. Genet. *2*, 173-179.

Olesen,C., Hansen,C., Bendsen,E., Byskov,A.G., Schwinger,E., Lopez-Pajares,I., Jensen,P.K., Kristoffersson,U., Schubert,R., Van Assche,E., Wahlstroem,J.,

Lespinasse,J., and Tommerup,N. (2001). Identification of human candidate genes for male infertility by digital differential display. Mol. Hum. Reprod. *7*, 11-20.

Paraoan,L., Grierson,I., and Maden,B.E. (2000). Analysis of expressed sequence tags of retinal pigment epithelium: cystatin C is an abundant transcript. Int. J. Biochem. Cell Biol. *32*, 417-426.

Paton,B.C., Schmid,B., Kustermann-Kuhn,B., Poulos,A., and Harzer,K. (1992). Additional biochemical findings in a patient and fetal sibling with a genetic defect in the sphingolipid activator protein (SAP) precursor, prosaposin. Evidence for a deficiency in SAP-1 and for a normal lysosomal neuraminidase. Biochem. J. *285 (Pt 2)*, 481-488.

Piatigorsky,J., Norman,B., Dishaw,L.J., Kos,L., Horwitz,J., Steinbach,P.J., and Kozmik,Z. (2001). J3-crystallin of the jellyfish lens: similarity to saposins. Proc. Natl. Acad. Sci. U. S A *98*, 12362-12367.

Piatigorsky,J., O'Brien,W.E., Norman,B.L., Kalumuck,K., Wistow,G.J., Borras,T., Nickerson,J.M., and Wawrousek,E.F. (1988). Gene sharing by delta-crystallin and argininosuccinate lyase. Proc. Natl. Acad. Sci. U. S. A *85*, 3479-3483.

Piatigorsky,J. and Wistow,G.J. (1989). Enzyme/crystallins: gene sharing as an evolutionary strategy. Cell *57*, 197-199.

Pontius,J.U., Wagner,L., and Schuler,G.D. (2003). UniGene: A Unified View of the Transcriptome. In The NCBI Handbook.

Pras,E., Frydman,M., Levy-Nissenbaum,E., Bakhan,T., Raz,J., Assia,E.I., Goldman,B., and Pras,E. (2000). A nonsense mutation (W9X) in CRYAA causes autosomal recessive cataract in an inbred Jewish Persian family. Invest Ophthalmol. Vis. Sci. *41*, 3511-3515.

Prosser,J. and van,H., V (1998). PAX6 mutations reviewed. Hum. Mutat. *11*, 93-108.

Putney,S.D., Herlihy,W.C., and Schimmel,P. (1983). A new troponin T and cDNA clones for 13 different muscle proteins, found by shotgun sequencing. Nature *302*, 718-721.

Quackenbush,J., Liang,F., Holt,I., Pertea,G., and Upton,J. (2000). The TIGR gene indices: reconstruction and representation of expressed gene sequences. Nucleic Acids Res. *28*, 141-145.

Quiring,R., Walldorf,U., Kloter,U., and Gehring,W.J. (1994). Homology of the eyeless gene of Drosophila to the Small eye gene in mice and Aniridia in humans. Science *265*, 785-789.

Rafi,M.A., de Gala,G., Zhang,X.L., and Wenger,D.A. (1993). Mutational analysis in a patient with a variant form of Gaucher disease caused by SAP-2 deficiency. Somat. Cell Mol. Genet. *19*, 1-7.

Ranta,S., Zhang,Y., Ross,B., Lonka,L., Takkunen,E., Messer,A., Sharp,J., Wheeler,R., Kusumi,K., Mole,S., Liu,W., Soares,M.B., Bonaldo,M.F., Hirvasniemi,A., de la,C.A., Gilliam,T.C., and Lehesjoki,A.E. (1999). The neuronal ceroid lipofuscinoses in human EPMR and mnd mutant mice are associated with mutations in CLN8. Nat. Genet. *23*, 233-236.

Rao,P.V., Huang,Q.L., Horwitz,J., and Zigler,J.S., Jr. (1995). Evidence that alpha-crystallin prevents non-specific protein aggregation in the intact eye lens. Biochim. Biophys. Acta *1245*, 439-447.

Rattner,A., Sun,H., and Nathans,J. (1999). Molecular genetics of human retinal disease. Annu. Rev. Genet. *33*, 89-131.

Raviola,G. and Raviola,E. (1978). Intercellular junctions in the ciliary epithelium. Invest Ophthalmol. Vis. Sci. *17*, 958-981.

Rezgaoui,M., Hermey,G., Riedel,I.B., Hampe,W., Schaller,H.C., and Hermans-Borgmeyer,I. (2001). Identification of SorCS2, a novel member of the VPS10 domain containing receptor family, prominently expressed in the developing mouse brain. Mech. Dev. *100*, 335-338.

Rodriguez,I.R. and Chader,G.J. (1992). A novel method for the isolation of tissue-specific genes. Nucleic Acids Res. *20*, 3528.

Roepman,R., Bernoud-Hubac,N., Schick,D.E., Maugeri,A., Berger,W., Ropers,H.H., Cremers,F.P., and Ferreira,P.A. (2000). The retinitis pigmentosa GTPase regulator (RPGR) interacts with novel transport-like proteins in the outer segments of rod photoreceptors. Hum. Mol. Genet. *9*, 2095-2105.

Rorman,E.G., Scheinker,V., and Grabowski,G.A. (1992). Structure and evolution of the human prosaposin chromosomal gene. Genomics *13*, 312-318.

Schena,M., Shalon,D., Davis,R.W., and Brown,P.O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science *270*, 467-470.

Scheurle,D., DeYoung,M.P., Binninger,D.M., Page,H., Jahanzeb,M., and Narayanan,R. (2000). Cancer gene discovery using digital differential display. Cancer Res *60*, 4037-4043.

Schnieders,F., Dork,T., Arnemann,J., Vogel,T., Werner,M., and Schmidtke,J. (1996). Testis-specific protein, Y-encoded (TSPY) expression in testicular tissues. Hum. Mol. Genet. *5*, 1801-1807.

Schoen,T.J., Mazuruk,K., Chader,G.J., and Rodriguez,I.R. (1995). Isolation of candidate genes for macular degeneration using an improved solid-phase subtractive cloning technique. Biochem. Biophys. Res Commun. *213*, 181-188.

Schuler,G.D. (1997). Pieces of the puzzle: expressed sequence tags and the catalog of human genes. J. Mol. Med. *75*, 694-698.

Schuler,G.D., Boguski,M.S., Stewart,E.A., Stein,L.D., Gyapay,G., Rice,K., White,R.E., Rodriguez-Tome,P., Aggarwal,A., Bajorek,E., Bentolila,S., Birren,B.B., Butler,A., Castle,A.B., Chiannilkulchai,N., Chu,A., Clee,C., Cowles,S., Day,P.J., Dibling,T., Drouot,N., Dunham,I., Duprat,S., East,C., Hudson,T.J., and . (1996). A gene map of the human genome. Science *274*, 540-546.

Sharma,S., Chang,J.T., Della,N.G., Campochiaro,P.A., and Zack,D.J. (2002). Identification of novel bovine RPE and retinal genes by subtractive hybridization. Mol. Vis. *8*, 251-258.

Sharon,D., Blackshaw,S., Cepko,C.L., and Dryja,T.P. (2002). Profile of the genes expressed in the human peripheral retina, macula, and retinal pigment epithelium determined through serial analysis of gene expression (SAGE). Proc. Natl. Acad. Sci. U. S. A *99*, 315-320.

Sharon,D., Sandberg,M.A., Rabe,V.W., Stillberger,M., Dryja,T.P., and Berson,E.L. (2003). RP2 and RPGR Mutations and Clinical Correlations in Patients with X-Linked Retinitis Pigmentosa. Am. J. Hum. Genet. *73*, 1131-1146.

Shimizu-Matsumoto,A., Adachi,W., Mizuno,K., Inazawa,J., Nishida,K., Kinoshita,S., Matsubara,K., and Okubo,K. (1997). An expression profile of genes in human retina and isolation of a complementary DNA for a novel rod photoreceptor protein. Invest Ophthalmol. Vis. Sci. *38*, 2576-2585.

Shoemaker,D.D., Schadt,E.E., Armour,C.D., He,Y.D., Garrett-Engele,P., McDonagh,P.D., Loerch,P.M., Leonardson,A., Lum,P.Y., Cavet,G., Wu,L.F., Altschuler,S.J., Edwards,S., King,J., Tsang,J.S., Schimmack,G., Schelter,J.M., Koch,J., Ziman,M., Marton,M.J., Li,B., Cundiff,P., Ward,T., Castle,J., Krolewski,M., Meyer,M.R., Mao,M., Burchard,J., Kidd,M.J., Dai,H., Phillips,J.W., Linsley,P.S., Stoughton,R., Scherer,S., and Boguski,M.S. (2001). Experimental annotation of the human genome using microarray technology. Nature *409*, 922-927.
Shimizu-Matsumoto,A., Adachi,W., Mizuno,K., Inazawa,J., Nishida,K., Kinoshita,S., Matsubara,K., and Okubo,K. (1997). An expression profile of genes in human retina and isolation of a complementary DNA for a novel rod photoreceptor protein. Invest Ophthalmol. Vis. Sci. *38*, 2576-2585.

Sikela,J.M. and Auffray,C. (1993). Finding new genes faster than ever. Nat. Genet. *3*, 189-191.

Simeone,A., Acampora,D., Gulisano,M., Stornaiuolo,A., and Boncinelli,E. (1992). Nested expression domains of four homeobox genes in developing rostral brain. Nature *358*, 687-690.

Simpson,T.I. and Price,D.J. (2002). Pax6; a pleiotropic player in development. Bioessays *24*, 1041-1051.

Singh,S., Mishra,R., Arango,N.A., Deng,J.M., Behringer,R.R., and Saunders,G.F. (2002). Iris hypoplasia in mice that lack the alternatively spliced Pax6(5a) isoform. Proc. Natl. Acad. Sci. U. S. A *99*, 6812-6815.

Sinha,S., Sharma,A., Agarwal,N., Swaroop,A., and Yang-Feng,T.L. (2000). Expression profile and chromosomal location of cDNA clones, identified from an enriched adult retina library. Invest Ophthalmol. Vis. Sci. *41*, 24-28.

Small,K.W., Weber,J.L., Roses,A., Lennon,F., Vance,J.M., and Pericak-Vance,M.A. (1992). North Carolina macular dystrophy is assigned to chromosome 6. Genomics *13*, 681-685.

Smith,L.M., Sanders,J.Z., Kaiser,R.J., Hughes,P., Dodd,C., Connell,C.R., Heiner,C., Kent,S.B., and Hood,L.E. (1986). Fluorescence detection in automated DNA sequence analysis. Nature *321*, 674-679.

Soares,M.B., Bonaldo,M.F., Jelene,P., Su,L., Lawton,L., and Efstratiadis,A. (1994). Construction and characterization of a normalized cDNA library. Proc. Natl. Acad. Sci. U. S. A *91*, 9228-9232.

Sohocki,M.M., Bowne,S.J., Sullivan,L.S., Blackshaw,S., Cepko,C.L., Payne,A.M., Bhattacharya,S.S., Khaliq,S., Qasim,M.S., Birch,D.G., Harrison,W.R., Elder,F.F., Heckenlively,J.R., and Daiger,S.P. (2000). Mutations in a new photoreceptor-pineal gene on 17p cause Leber congenital amaurosis. Nat. Genet. *24*, 79-83.

Sohocki,M.M., Malone,K.A., Sullivan,L.S., and Daiger,S.P. (1999). Localization of retina/pineal-expressed sequences: identification of novel candidate genes for inherited retinal disorders. Genomics *58*, 29-33.

Sosnowski,R.G., Tu,E., Butler,W.F., O'Connell,J.P., and Heller,M.J. (1997). Rapid determination of single base mismatch mutations in DNA hybrids by direct electric field control. Proc. Natl. Acad. Sci. U. S. A *94*, 1119-1123.

Spirin,K.S., Ljubimov,A.V., Castellon,R., Wiedoeft,O., Marano,M., Sheppard,D., Kenney,M.C., and Brown,D.J. (1999). Analysis of gene expression in human bullous keratopathy corneas containing limiting amounts of RNA. Invest Ophthalmol. Vis. Sci. *40*, 3108-3115.

Steel,K.P., Davidson,D.R., and Jackson,I.J. (1992). TRP-2/DT, a new early melanoblast marker, shows that steel growth factor (c-kit ligand) is a survival factor. Development *115*, 1111-1119.

Stohr,H., Mah,N., Schulz,H., Gehrig,A., Frohlich,S., and Weber,B. (2000). EST mining of the UniGene dataset to identify retina-specific genes. Cytogenet. Cell Genet. *91*, 267-277.

Stoykova,A. and Gruss,P. (1994). Roles of Pax-genes in developing and adult brain as suggested by expression patterns. J. Neurosci. *14*, 1395-1412.

Sun,J.K., Iwata,T., Zigler,J.S., Jr., and Carper,D.A. (2000). Differential gene expression in male and female rat lenses undergoing cataract induction by transforming growth factor-beta (TGF-beta). Exp. Eye Res *70*, 169-181.

Sun,Y., Witte,D.P., and Grabowski,G.A. (1994). Developmental and tissue-specific expression of prosaposin mRNA in murine tissues. Am. J. Pathol. *145*, 1390-1398.

Sun,Y., Witte,D.P., Jin,P., and Grabowski,G.A. (2003). Analyses of temporal regulatory elements of the prosaposin gene in transgenic mice. Biochem. J. *370*, 557-566.

Swanson,D.A., Freund,C.L., Steel,J.M., Xu,S., Ploder,L., McInnes,R.R., and Valle,D. (1997). A differential hybridization scheme to identify photoreceptor-specific genes. Genome Res. *7*, 513-521.

Takahashi,N., Hashida,H., Zhao,N., Misumi,Y., and Sakaki,Y. (1995). High-density cDNA filter analysis of the expression profiles of the genes preferentially expressed in human brain. Gene *164*, 219-227.

Theiler,K., Varnum,D.S., and Stevens,L.C. (1978). Development of Dickie's small eye, a mutation in the house mouse. Anat. Embryol. (Berl) *155*, 81-86.

Thompson,J.D., Higgins,D.G., and Gibson,T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. *22*, 4673-4680.

Thut,C.J., Rountree,R.B., Hwa,M., and Kingsley,D.M. (2001). A large-scale in situ screen provides molecular evidence for the induction of eye anterior segment structures by the developing lens. Dev. Biol. *231*, 63-76.

Ton,C.C., Hirvonen,H., Miwa,H., Weil,M.M., Monaghan,P., Jordan,T., van,H., V, Hastie,N.D., Meijers-Heijboer,H., Drechsler,M., and . (1991). Positional cloning and characterization of a paired box- and homeobox-containing gene from the aniridia region. Cell *67*, 1059-1074.

Ton,C.C., Miwa,H., and Saunders,G.F. (1992). Small eye (Sey): cloning and characterization of the murine homolog of the human aniridia gene. Genomics *13*, 251-256.

Turque,N., Plaza,S., Radvanyi,F., Carriere,C., and Saule,S. (1994). Pax-QNR/Pax-6, a paired box- and homeobox-containing gene expressed in neurons, is also expressed in pancreatic endocrine cells. Mol. Endocrinol. *8*, 929-938.

Ullrich,A., Sures,I., D'Egidio,M., Jallal,B., Powell,T.J., Herbst,R., Dreps,A., Azam,M., Rubinstein,M., Natoli,C., and . (1994). The secreted tumor-associated antigen 90K is a potent immune stimulator. J. Biol. Chem. *269*, 18401-18407.

van der Spuy,J., Chapple,J.P., Clark,B.J., Luthert,P.J., Sethi,C.S., and Cheetham,M.E. (2002). The Leber congenital amaurosis gene product AIPL1 is localized exclusively in rod photoreceptors of the adult human retina. Hum. Mol. Genet. *11*, 823-831.

Van Leen,R.W., Breuer,M.L., Lubsen,N.H., and Schoenmakers,J.G. (1987). Developmental expression of crystallin genes: in situ hybridization reveals a differential localization of specific mRNAs. Dev. Biol. *123*, 338-345.

Velculescu,V.E., Zhang,L., Vogelstein,B., and Kinzler,K.W. (1995). Serial analysis of gene expression. Science *270*, 484-487.

Venter,J.C., Adams,M.D., Myers,E.W., Li,P.W., Mural,R.J., Sutton,G.G., Smith,H.O., Yandell,M., Evans,C.A., Holt,R.A., Gocayne,J.D., Amanatides,P., Ballew,R.M., Huson,D.H., Wortman,J.R., Zhang,Q., Kodira,C.D., Zheng,X.H., Chen,L., Skupski,M., Subramanian,G., Thomas,P.D., Zhang,J., Gabor Miklos,G.L., Nelson,C., Broder,S., Clark,A.G., Nadeau,J., McKusick,V.A., Zinder,N., Levine,A.J., Roberts,R.J., Simon,M., Slayman,C., Hunkapiller,M., Bolanos,R., Delcher,A., Dew,I., Fasulo,D., Flanigan,M., Florea,L., Halpern,A., Hannenhalli,S., Kravitz,S., Levy,S., Mobarry,C., Reinert,K., Remington,K., Abu-Threideh,J., Beasley,E., Biddick,K., Bonazzi,V., Brandon,R., Cargill,M., Chandramouliswaran,I., Charlab,R., Chaturvedi,K., Deng,Z., Di,F., V, Dunn,P., Eilbeck,K., Evangelista,C., Gabrielian,A.E., Gan,W., Ge,W., Gong,F., Gu,Z., Guan,P., Heiman,T.J., Higgins,M.E., Ji,R.R., Ke,Z., Ketchum,K.A., Lai,Z., Lei,Y., Li,Z., Li,J., Liang,Y., Lin,X., Lu,F., Merkulov,G.V., Milshina,N., Moore,H.M., Naik,A.K., Narayan,V.A., Neelam,B., Nusskern,D., Rusch,D.B., Salzberg,S., Shao,W., Shue,B., Sun,J., Wang,Z., Wang,A., Wang,X., Wang,J., Wei,M., Wides,R., Xiao,C., Yan,C., Yao,A., Ye,J., Zhan,M., Zhang,W., Zhang,H., Zhao,Q., Zheng,L., Zhong,F., Zhong,W., Zhu,S., Zhao,S., Gilbert,D., Baumhueter,S., Spier,G., Carter,C., Cravchik,A., Woodage,T., Ali,F., An,H., Awe,A., Baldwin,D., Baden,H., Barnstead,M., Barrow,I., Beeson,K., Busam,D., Carver,A., Center,A., Cheng,M.L., Curry,L., Danaher,S., Davenport,L., Desilets,R., Dietz,S., Dodson,K., Doup,L., Ferriera,S., Garg,N., Gluecksmann,A., Hart,B., Haynes,J., Haynes,C., Heiner,C., Hladun,S., Hostin,D., Houck,J., Howland,T., Ibegwam,C., Johnson,J., Kalush,F., Kline,L., Koduru,S., Love,A., Mann,F., May,D., McCawley,S., McIntosh,T., McMullen,I., Moy,M., Moy,L., Murphy,B., Nelson,K., Pfannkoch,C., Pratts,E., Puri,V., Qureshi,H., Reardon,M., Rodriguez,R., Rogers,Y.H., Romblad,D., Ruhfel,B., Scott,R., Sitter,C., Smallwood,M., Stewart,E., Strong,R., Suh,E., Thomas,R., Tint,N.N., Tse,S., Vech,C., Wang,G., Wetter,J., Williams,S., Williams,M., Windsor,S., Winn-Deen,E., Wolfe,K., Zaveri,J., Zaveri,K., Abril,J.F., Guigo,R., Campbell,M.J., Sjolander,K.V., Karlak,B., Kejariwal,A., Mi,H., Lazareva,B., Hatton,T., Narechania,A., Diemer,K., Muruganujan,A., Guo,N., Sato,S., Bafna,V., Istrail,S., Lippert,R., Schwartz,R., Walenz,B., Yooseph,S., Allen,D., Basu,A., Baxendale,J., Blick,L., Caminha,M., Carnes-Stine,J., Caulk,P., Chiang,Y.H., Coyne,M., Dahlke,C., Mays,A., Dombroski,M., Donnelly,M., Ely,D., Esparham,S., Fosler,C., Gire,H., Glanowski,S., Glasser,K., Glodek,A., Gorokhov,M., Graham,K., Gropman,B., Harris,M., Heil,J., Henderson,S., Hoover,J., Jennings,D., Jordan,C., Jordan,J., Kasha,J., Kagan,L., Kraft,C., Levitsky,A., Lewis,M., Liu,X., Lopez,J., Ma,D., Majoros,W., McDaniel,J., Murphy,S., Newman,M., Nguyen,T., Nguyen,N., and Nodell,M. (2001). The sequence of the human genome. Science *291*, 1304-1351.

Veres,G., Gibbs,R.A., Scherer,S.E., and Caskey,C.T. (1987). The molecular basis of the sparse fur mouse mutation. Science *237*, 415-417.

Vervoort,R., Lennon,A., Bird,A.C., Tulloch,B., Axton,R., Miano,M.G., Meindl,A., Meitinger,T., Ciccodicola,A., and Wright,A.F. (2000). Mutational hot spot within a new RPGR exon in X-linked retinitis pigmentosa. Nat. Genet. *25*, 462-466.

Vervoort,R. and Wright,A.F. (2002). Mutations of RPGR in X-linked retinitis pigmentosa (RP3). Hum. Mutat. *19*, 486-500.

Vicart,P., Caron,A., Guicheney,P., Li,Z., Prevost,M.C., Faure,A., Chateau,D., Chapon,F., Tome,F., Dupret,J.M., Paulin,D., and Fardeau,M. (1998). A missense mutation in the alphaB-crystallin chaperone gene causes a desmin-related myopathy. Nat. Genet. *20*, 92-95.

Vogel,T., Boettger-Tong,H., Nanda,I., Dechend,F., Agulnik,A.I., Bishop,C.E., Schmid,M., and Schmidtke,J. (1998). A murine TSPY. Chromosome. Res. *6*, 35-40.

Vogel,T. and Schmidtke,J. (1998). Structure and function of TSPY, the Y-chromosome gene coding for the "testis-specific protein". Cytogenet. Cell Genet. *80*, 209-213.

Wallis,D.E., Roessler,E., Hehr,U., Nanni,L., Wiltshire,T., Richieri-Costa,A., Gillessen-Kaesbach,G., Zackai,E.H., Rommens,J., and Muenke,M. (1999). Mutations in the homeodomain of the human SIX3 gene cause holoprosencephaly. Nat. Genet. *22*, 196-198.

Walt,D.R. (2000). Techview: molecular biology. Bead-based fiber-optic arrays. Science *287*, 451-452.

Walther,C. and Gruss,P. (1991). Pax-6, a murine paired box gene, is expressed in the developing CNS. Development *113*, 1435-1449.

Wang,K. and Spector,A. (1994). The chaperone activity of bovine alpha crystallin. Interaction with other lens crystallins in native and denatured states. J. Biol. Chem. *269*, 13601-13608.

Wang,Y., Macke,J.P., Abella,B.S., Andreasson,K., Worley,P., Gilbert,D.J., Copeland,N.G., Jenkins,N.A., and Nathans,J. (1996). A large family of putative transmembrane receptors homologous to the product of the Drosophila tissue polarity gene frizzled. J. Biol. Chem. *271*, 4468-4476.

Waterston,R.H., Lindblad-Toh,K., Birney,E., Rogers,J., Abril,J.F., Agarwal,P., Agarwala,R., Ainscough,R., Alexandersson,M., An,P., Antonarakis,S.E., Attwood,J., Baertsch,R., Bailey,J., Barlow,K., Beck,S., Berry,E., Birren,B., Bloom,T., Bork,P., Botcherby,M., Bray,N., Brent,M.R., Brown,D.G., Brown,S.D., Bult,C., Burton,J., Butler,J., Campbell,R.D., Carninci,P., Cawley,S., Chiaromonte,F., Chinwalla,A.T., Church,D.M., Clamp,M., Clee,C., Collins,F.S., Cook,L.L., Copley,R.R., Coulson,A., Couronne,O., Cuff,J., Curwen,V., Cutts,T., Daly,M., David,R., Davies,J., Delehaunty,K.D., Deri,J., Dermitzakis,E.T., Dewey,C., Dickens,N.J., Diekhans,M., Dodge,S., Dubchak,I., Dunn,D.M., Eddy,S.R., Elnitski,L., Emes,R.D., Eswara,P., Eyras,E., Felsenfeld,A., Fewell,G.A., Flicek,P., Foley,K., Frankel,W.N., Fulton,L.A.,

Fulton,R.S., Furey,T.S., Gage,D., Gibbs,R.A., Glusman,G., Gnerre,S., Goldman,N., Goodstadt,L., Grafham,D., Graves,T.A., Green,E.D., Gregory,S., Guigo,R., Guyer,M., Hardison,R.C., Haussler,D., Hayashizaki,Y., Hillier,L.W., Hinrichs,A., Hlavina,W., Holzer,T., Hsu,F., Hua,A., Hubbard,T., Hunt,A., Jackson,I., Jaffe,D.B., Johnson,L.S., Jones,M., Jones,T.A., Joy,A., Kamal,M., Karlsson,E.K., Karolchik,D., Kasprzyk,A., Kawai,J., Keibler,E., Kells,C., Kent,W.J., Kirby,A., Kolbe,D.L., Korf,I., Kucherlapati,R.S., Kulbokas,E.J., Kulp,D., Landers,T., Leger,J.P., Leonard,S., Letunic,I., Levine,R., Li,J., Li,M., Lloyd,C., Lucas,S., Ma,B., Maglott,D.R., Mardis,E.R., Matthews,L., Mauceli,E., Mayer,J.H., McCarthy,M., McCombie,W.R., McLaren,S., McLay,K., McPherson,J.D., Meldrim,J., Meredith,B., Mesirov,J.P., Miller,W., Miner,T.L., Mongin,E., Montgomery,K.T., Morgan,M., Mott,R., Mullikin,J.C., Muzny,D.M., Nash,W.E., Nelson,J.O., Nhan,M.N., Nicol,R., Ning,Z., Nusbaum,C., O'Connor,M.J., Okazaki,Y., Oliver,K., Overton-Larty,E., Pachter,L., Parra,G., Pepin,K.H., Peterson,J., Pevzner,P., Plumb,R., Pohl,C.S., Poliakov,A., Ponce,T.C., Ponting,C.P., Potter,S., Quail,M., Reymond,A., Roe,B.A., Roskin,K.M., Rubin,E.M., Rust,A.G., Santos,R., Sapojnikov,V., Schultz,B., Schultz,J., Schwartz,M.S., Schwartz,S., Scott,C., Seaman,S., Searle,S., Sharpe,T., Sheridan,A., Shownkeen,R., Sims,S., Singer,J.B., Slater,G., Smit,A., Smith,D.R., Spencer,B., Stabenau,A., Stange-Thomann,N., Sugnet,C., Suyama,M., Tesler,G., Thompson,J., Torrents,D., Trevaskis,E., Tromp,J., Ucla,C., Ureta-Vidal,A., Vinson,J.P., Von Niederhausern,A.C., Wade,C.M., Wall,M., Weber,R.J., Weiss,R.B., Wendl,M.C., West,A.P., Wetterstrand,K., Wheeler,R., Whelan,S., Wierzbowski,J., Willey,D., Williams,S., Wilson,R.K., Winter,E., Worley,K.C., Wyman,D., Yang,S., Yang,S.P., Zdobnov,E.M., Zody,M.C., and Lander,E.S. (2002). Initial sequencing and comparative analysis of the mouse genome. Nature *420*, 520-562.

Watson,A.J., Katz,A., and Simon,M.I. (1994). A fifth member of the mammalian G-protein beta-subunit family. Expression in brain and activation of the beta 2 isotype of phospholipase C. J. Biol. Chem. *269*, 22150-22156.

Weeks,D.E., Conley,Y.P., Tsai,H.J., Mah,T.S., Rosenfeld,P.J., Paul,T.O., Eller,A.W., Morse,L.S., Dailey,J.P., Ferrell,R.E., and Gorin,M.B. (2001). Age-related maculopathy: an expanded genome-wide scan with evidence of susceptibility loci within the 1q31 and 17q25 regions. Am. J. Ophthalmol. *132*, 682-692.

Welsh,J., Chada,K., Dalal,S.S., Cheng,R., Ralph,D., and McClelland,M. (1992). Arbitrarily primed PCR fingerprinting of RNA. Nucleic Acids Res. *20*, 4965-4970.

Wheeler,D.L., Church,D.M., Federhen,S., Lash,A.E., Madden,T.L., Pontius,J.U., Schuler,G.D., Schriml,L.M., Sequeira,E., Tatusova,T.A., and Wagner,L. (2003). Database resources of the National Center for Biotechnology. Nucleic Acids Res. *31*, 28-33.

Whitaker,T.L., Steinbrecher,K.A., Copeland,N.G., Gilbert,D.J., Jenkins,N.A., and Cohen,M.B. (1997). The uroguanylin gene (Guca1b) is linked to guanylin (Guca2) on mouse chromosome 4. Genomics *45*, 348-354.

Winter,E. and Ponting,C.P. (2002). TRAM, LAG1 and CLN8: members of a novel family of lipid-sensing domains? Trends Biochem. Sci. *27*, 381-383.

Wistow,G. (2002). A project for ocular bioinformatics: NEIBank. Mol. Vis. *8*, 161-163.

Wistow,G., Bernstein,S.L., Wyatt,M.K., Fariss,R.N., Behal,A., Touchman,J.W., Bouffard,G., Smith,D., and Peterson,K. (2002a). Expressed sequence tag analysis of human RPE/choroid for the NEIBank Project: over 6000 non-redundant transcripts, novel genes and splice variants. Mol. Vis. *8*, 205-220.

Wistow,G., Berstein,S.L., Ray,S., Wyatt,M.K., Behal,A., Touchman,J.W., Bouffard,G., Smith,D., and Peterson,K. (2002b). Expressed sequence tag analysis of adult human iris for the NEIBank Project: steroid-response factors and similarities with retinal pigment epithelium. Mol. Vis. *8*, 185-195.

Wistow,G., Berstein,S.L., Wyatt,M.K., Behal,A., Touchman,J.W., Bouffard,G., Smith,D., and Peterson,K. (2002c). Expressed sequence tag analysis of adult human lens for the NEIBank Project: over 2000 non-redundant transcripts, novel genes and splice variants. Mol. Vis. *8*, 171-184.

Wistow,G., Berstein,S.L., Wyatt,M.K., Ray,S., Behal,A., Touchman,J.W., Bouffard,G., Smith,D., and Peterson,K. (2002d). Expressed sequence tag analysis of human retina for the NEIBank Project: retbindin, an abundant, novel retinal cDNA and alternative splicing of other retina-preferred gene transcripts. Mol. Vis. *8*, 196-204.

Wistow,G.J., Lietman,T., Williams,L.A., Stapel,S.O., de Jong,W.W., Horwitz,J., and Piatigorsky,J. (1988). Tau-crystallin/alpha-enolase: one gene encodes both an enzyme and a lens structural protein. J. Cell Biol. *107*, 2729-2736.

Wistow,G.J. and Piatigorsky,J. (1988). Lens crystallins: the evolution and expression of proteins for a highly specialized tissue. Annu. Rev. Biochem. *57*, 479-504.

Wong,C.W. and Privalsky,M.L. (1998). Components of the SMRT corepressor complex exhibit distinctive interactions with the POZ domain oncoproteins PLZF, PLZF-RARalpha, and BCL-6. J. Biol. Chem. *273*, 27695-27702.

Yang,J.H., Gross,R.L., Basinger,S.F., and Wu,S.M. (2001). Apoptotic cell death of cultured salamander photoreceptors induced by cccp: CsA-insensitive mitochondrial permeability transition. J. Cell Sci. *114*, 1655-1664.

Yelin,R., Dahary,D., Sorek,R., Levanon,E.Y., Goldstein,O., Shoshan,A., Diber,A., Biton,S., Tamir,Y., Khosravi,R., Nemzer,S., Pinner,E., Walach,S., Bernstein,J., Savitsky,K., and Rotman,G. (2003). Widespread occurrence of antisense transcription in the human genome. Nat. Biotechnol. *21*, 379-386.

Yokoyama,A., Maruiwa,F., Hayakawa,M., Kanai,A., Vervoort,R., Wright,A.F., Yamada,K., Niikawa,N., and Naoi,N. (2001). Three novel mutations of the RPGR gene exon ORF15 in three Japanese families with X-linked retinitis pigmentosa. Am. J. Med. Genet. *104*, 232-238.

Yoshida,S., Yashar,B.M., Hiriyanna,S., and Swaroop,A. (2002). Microarray analysis of gene expression in the aging human retina. Invest Ophthalmol. Vis. Sci. *43*, 2554-2560.

Zhang,J.S., Yang-Feng,T.L., Muller,U., Mohandas,T.K., de Jong,P.J., and Lau,Y.F. (1992). Molecular isolation and characterization of an expressed gene from the human Y chromosome. Hum. Mol. Genet. *1*, 717-726.

Zhao,N., Hashida,H., Takahashi,N., Misumi,Y., and Sakaki,Y. (1995). High-density cDNA filter analysis: a novel approach for large-scale, quantitative analysis of gene expression. Gene *156*, 207-213.

Zhou,R.H., Kokame,K., Tsukamoto,Y., Yutani,C., Kato,H., and Miyata,T. (2001). Characterization of the human NDRG gene family: a newly identified member, NDRG4, is specifically expressed in brain and heart. Genomics *73*, 86-97.

Zhu,C.C., Dyer,M.A., Uchikawa,M., Kondoh,H., Lagutin,O.V., and Oliver,G. (2002). Six3-mediated auto repression and eye development requires its interaction with members of the Groucho-related family of co-repressors. Development *129*, 2835-2849.

Zollman,S., Godt,D., Prive,G.G., Couderc,J.L., and Laski,F.A. (1994). The BTB domain, found primarily in zinc finger proteins, defines an evolutionarily conserved family that includes several developmentally regulated genes in Drosophila. Proc. Natl. Acad. Sci. U. S. A *91*, 10717-10721.

Appendices

Appendix A

Known genes in mouse and human DDD results

Eye relevant	Mm Symbol	Mm Name	Hs Symbol	Hs Name	Mm build 82	Mm build 82: recent analysis	Mm build 100	Mm build 100: recent analysis	Mm build 120	Hs build 160
Y	*Abca4*	ATP-binding cassette, sub-family A (ABC1), member 4	*ABCA4*	ATP-binding cassette, sub-family A (ABC1), member 4			1		1	1
Y	*Abcg4*	ATP-binding cassette, sub-family G (WHITE), member 4							1	
N	*Actg*	actin, gamma, cytoplasmic							1	
N			*ADD1*	adducin 1 (alpha)						1
Y	*Aipl1*	aryl hydrocarbon receptor-interacting protein-like 1	*AIPL1*	aryl hydrocarbon receptor interacting protein-like 1		2		1	1	1
Y	*Aldh1a1*	aldehyde dehydrogenase family 1, subfamily A1							1	
N	*Aldo1*	aldolase 1, A isoform	*ALDOA*	aldolase A, fructose-bisphosphate					1	1
N			*ALDOC*	aldolase C, fructose-bisphosphate						1
N	*Als2*	amyotrophic lateral sclerosis 2 (juvenile) homolog (human)	*ALS2CR5*	Sapiens ALS2CR5 mRNA, splicing variant					1	1
N			*APP*	amyloid beta (A4) precursor protein (protease nexin-II, Alzheimer disease)						1
N			*ARF4L*	ADP-ribosylation factor 4-like						1
Y	*Arr3*	arrestin 3, retinal	*ARR3*	arrestin 3, retinal (X-arrestin)			1		1	1
N	*Arhg*	ras homolog gene family, member G				1				
N	*Asf1a*	ASF1 anti-silencing function 1 homolog A (S. cerevisiae)				1				
Y	*Atoh7*	atonal homolog 7 (Drosophila)					1		1	
N	*Atp1a3*	ATPase, Na+/K+ transporting, alpha 3 polypeptide	*ATP1A3*	ATPase, Na+/K+ transporting, alpha 3 polypeptide				1		1
Y	*Atp1b2*	ATPase, Na+/K+ transporting, beta 2 polypeptide	*ATP1B2*	ATPase, Na+/K+ transporting, beta 2 polypeptide			1		1	1
N			*B3GAT2*	beta-1,3-glucuronyltransferase 2 (glucuronosyltransferase S)						1
Y	*Barhl2*	BarH-like 2 (Drosophila)							1	
N	*Baf1a*	HLA-B-associated transcript 1A							1	
N	*Bcat1*	branched chain aminotransferase 1, cytosolic					1		1	
Y	*Bfsp1*	beaded-filament structure protein	*BFSP1*	beaded filament structural protein 1, filensin		1	1		1	1
Y			*BFSP2*	beaded filament structural protein 2, phakinin						1
N	*Bin1*	bridging integrator 1					1		1	
Y			*BIRC7*	baculoviral IAP repeat-containing 7 (livin)						1
N	*Bup*	Bmi1 upstream gene				1				
N			*C4A*	complement component 4A						1
Y			*C7orf9*	chromosome 7 open reading frame 9						1
N			*C7orf20*	chromosome 7 open reading frame 20						1
N			*C12orf3*	chromosome 12 open reading frame 3						1
Y			*C18orf2*	chromosome 18 open reading frame 2						1
Y	*Cabp5*	calcium binding protein 5	*CABP5*	calcium binding protein 5			1		1	1
Y	*Cacna1f*	calcium channel, voltage-dependent, alpha 1F subunit	*CACNA1F*	calcium channel, voltage-dependent, alpha 1F subunit			1		1	1
N	*Cacna2d3*	calcium channel, voltage dependent, alpha2/delta subunit 3					1			
N	*Cacng4*	calcium channel, voltage-dependent, gamma subunit 4					1		1	
Y	*Capn3*	calpain 3			1		1	1	1	
N	*Cart*	cocaine and amphetamine regulated transcript							1	
N	*Casq2*	calsequestrin 2	*CASQ2*	calsequestrin 2 (cardiac muscle)	1	1	1		1	1
N	*Cbln2*	cerebellin 2 precursor protein							1	
Y	*Ccnd1*	cyclin D1			1		1		1	
N	*Ccnl*	cyclin L1						1	1	
N			*CCNI*	cyclin I						1
N			*CDC42BPB*	CDC42 binding protein kinase beta (DMPK-like)						1
Y			*CDH23*	cadherin related 23						1
Y			*CDT6*	angiopoietin-like factor						1
N			*CGI-41*	CGI-41 protein						1
Y			*CHRNA3*	cholinergic receptor, nicotinic, alpha polypeptide 3						1
Y	*Chrna6*	cholinergic receptor, nicotinic, alpha polypeptide 6				1	1		1	
N	*Chst10*	carbohydrate sulfotransferase 10				1				
Y	*Chx10*	C. elegans ceh-10 homeo domain containing homolog					1		1	
N	*Ckb*	creatine kinase, brain	*CKB*	creatine kinase, brain			1		1	1
Y			*CLCN7*	chloride channel 7						1
N	*Cldn2*	claudin 2						1		
N	*Clk*	CDC-like kinase							1	
N	*Clstn1*	calsyntenin 1					1		1	
N			*CLU*	clusterin (complement lysis inhibitor, SP-40,40, sulfated glycoprotein 2, testosterone-repressed prostate message 2, apolipoprotein J)						1
N	*Cmas*	cytidine monophospho-N-acetylneuraminic acid synthetase			1				1	
Y	*Cnga1*	cyclic nucleotide gated channel alpha 1					1		1	
Y			*CNGB1*	cyclic nucleotide gated channel beta 1						1
N	*Cnn3*	calponin 3, acidic					1			
N	*Col1a2*	procollagen, type I, alpha 2			1					
Y	*Col9a1*	procollagen, type IX, alpha 1			1				1	
Y			*CP*	ceruloplasmin (ferroxidase)						1
N	*Cpe*	carboxypeptidase E				1				
N	*Cplx4*	complexin IV						1		
N			*CRABP1*	cellular retinoic acid binding protein 1						1
Y	*Crb1*	crumbs homolog 1 (Drosophila)							1	
N	*Crmp1*	collapsin response mediator protein 1	*CRMP1*	collapsin response mediator protein 1			1			1
N			*CRTAC1*	cartilage acidic protein 1						1
Y	*Crx*	cone-rod homeobox containing gene	*CRX*	cone-rod homeobox			1		1	1
Y	*Cryaa*	crystallin, alpha A	*CRYAA*	crystallin, alpha A	1		1		1	1
Y			*CRYAB*	crystallin, alpha B						1
Y	*Cryba1*	crystallin, beta A1	*CRYBA1*	crystallin, beta A1	1		1		1	1
Y	*Cryba2*	crystallin, beta A2	*CRYBA2*	crystallin, beta A2	1		1		1	1
Y	*Cryba4*	crystallin, beta A4	*CRYBA4*	crystallin, beta A4					1	1
Y	*Crybb1*	crystallin, beta B1	*CRYBB1*	crystallin, beta B1	1		1		1	1

Eye relevant	Mm Symbol	Mm Name	Hs Symbol	Hs Name	Mm build 82	Mm build 82: recent analysis	Mm build 100	Mm build 100: recent analysis	Mm build 120	Hs build 160
Y	*Crybb2*	crystallin, beta B2	*CRYBB2*	crystallin, beta B2	1	1	1		1	1
Y	*Crybb3*	crystallin, beta B3	*CRYBB3*	crystallin, beta B3	1		1		1	1
Y	*Cryga*	crystallin, gamma A			1		1		1	
Y	*Crygb*	crystallin, gamma B			1		1		1	
Y	*Crygc*	crystallin, gamma C	*CRYGC*	crystallin, gamma C	1		1		1	1
Y	*Crygd*	crystallin, gamma D	*CRYGD*	crystallin, gamma D	1		1		1	1
Y	*Cryge*	crystallin, gamma E							1	
Y	*Crygf*	crystallin, gamma F			1		1	1	1	
Y	*Crygn*	crystallin, gamma N							1	
Y	*Crygs*	crystallin, gamma S	*CRYGS*	crystallin, gamma S					1	1
N	*Ctbp2*	C-terminal binding protein 2							1	
N	*Ctsf*	cathepsin F	*CTSF*	cathepsin F			1		1	1
N			*CXCL14*	chemokine (C-X-C motif) ligand 14						1
N	*D5Bwg0860e*	DNA segment, Chr 5, Brigham & Women's Genetics 0860 expressed						1		
N	*D15Ertd806e*	DNA segment, Chr 15, ERATO Doi 806, expressed				1		1		
Y	*Dct*	dopachrome tautomerase			1	1	1		1	
N	*Ddr1*	discoidin domain receptor family, member 1			1					
N	*Ddx3*	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 3							1	
N	*Ddx5*	DEAD (Asp-Glu-Ala-Asp) box polypeptide 5							1	
N			*DDX51*	DEAD (Asp-Glu-Ala-Asp) box polypeptide 51						1
N	*DDXx*	nucleic acid helicase DDXx				1		1		
N			*DIRAS1*	DIRAS family, GTP-binding RAS-like 1						1
N	*Dkk3*	dickkopf homolog 3 (Xenopus laevis)			1		1		1	
N	*Dlgh4*	discs, large homolog 4 (Drosophila)					1		1	
N	*Dnm*	dynamin					1		1	
N	*Doc2b*	double C2, beta					1			
N	*Dpf1l*	deleted in polyposis 1-like-1							1	
N	*Dpysl3*	dihydropyrimidinase-like 3					1			
Y	*Drd4*	dopamine receptor 4			1		1		1	
N			*DUSP1*	dual specificity phosphatase 1						1
N	*Edn2*	endothelin 2			1		1		1	
N	*Eef1a1*	eukaryotic translation elongation factor 1 alpha 1					1		1	
N			*EEF1A2*	eukaryotic translation elongation factor 1 alpha 2						1
N			*EGFL3*	EGF-like-domain, multiple 3						1
Y	*Elovl4*	elongation of very long chain fatty acids (FEN1/Elo2, SUR4/Elo3, yeast)-like 4					1		1	
N	*Elovl5*	ELOVL family member 5, elongation of long chain fatty acids (yeast)				1		1		
N	*Eno1*	enolase 1, alpha non-neuron					1		1	
N	*Eno2*	enolase 2, gamma neuronal	*ENO2*	enolase 2, (gamma, neuronal)			1			1
N	*Epb4.1*	erythrocyte protein band 4.1							1	
N	*Epb4.1l2*	erythrocyte protein band 4.1-like 2					1		1	
N	*Eraf*	erythroid associated factor							1	
N	*Espn*	espin			1		1		1	
N			*ESRRG*	estrogen-related receptor gamma						1
N			*ETFB*	electron-transfer-flavoprotein, beta polypeptide						1
Y	*Eya2*	eyes absent 2 homolog (Drosophila)			1		1		1	
N	*Fancc*	Fanconi anemia, complementation group C						1		
N	*Fgf15*	fibroblast growth factor 15					1		1	
N	*Fmip*	Fms interacting protein							1	
N	*Foxm1*	forkhead box M1							1	
N			*FRZB*	frizzled-related protein						1
N			*FSCN1*	fascin homolog 1, actin-bundling protein (Strongylocentrotus purpuratus)						1
Y	*Fscn2*	fascin homolog 2, actin-bundling protein, retinal (Strongylocentrotus purpuratus)	*FSCN2*	fascin homolog 2, actin-bundling protein, retinal (Strongylocentrotus purpuratus)				1		1
N	*Fus*	fusion, derived from t(12;16) malignant liposarcoma (human)					1			
N			*FXYD1*	FXYD domain containing ion transport regulator 1 (phospholemman						1
N	*Gab3*	growth factor receptor bound protein 2-associated protein 3							1	
N	*Gabt4*	gamma-aminobutyric acid (GABA-A) transporter 4				1		1	1	
N	*Gap43*	growth associated protein 43					1			
N	*Gapd*	glyceraldehyde-3-phosphate dehydrogenase	*GAPD*	glyceraldehyde-3-phosphate dehydrogenase			1		1	1
N			*GDF5*	growth differentiation factor 5 (cartilage-derived morphogenetic protein-1)						1
N	*Gldc*	glycine decarboxylase			1		1		1	
N			*GLUL*	glutamate-ammonia ligase (glutamine synthase)						1
N			*GLULD1*	glutamate-ammonia ligase (glutamine synthase) domain containing 1						1
Y	*Gnat1*	guanine nucleotide binding protein, alpha transducing 1	*GNAT1*	guanine nucleotide binding protein (G protein), alpha transducing activity polypeptide 1			1		1	1
Y	*Gnat2*	guanine nucleotide binding protein, alpha transducing 2					1		1	
Y	*Gnb1*	guanine nucleotide binding protein, beta 1					1		1	
N	*Gnb3*	guanine nucleotide binding protein, beta 3	*GNB3*	guanine nucleotide binding protein (G protein), beta polypeptide 3			1		1	1
N	*Gnb5*	guanine nucleotide binding protein, beta 5	*GNB5*	guanine nucleotide binding protein (G protein), beta 5			1			1
N	*Gng1*	guanine nucleotide binding protein (G protein), gamma 1 subunit				1	1		1	
Y			*GNGT1*	guanine nucleotide binding protein (G protein), gamma transducing activity polypeptide 1						1
N	*Gpi1*	glucose phosphate isomerase 1							1	
N	*Gpm6a*	glycoprotein m6a						1		
N			*GPX3*	glutathione peroxidase 3 (plasma)						1
Y	*Grifin*	galectin-related inter-fiber protein							1	
N			*GRIN1*	glutamate receptor, ionotropic, N-methyl D-aspartate 1						1
Y	*Grm6*	glutamate receptor, metabotropic 6							1	
N	*Grtp1*	GH regulated TBC protein 1				1		1		

Eye relevant	Mm Symbol	Mm Name	Hs Symbol	Hs Name	Mm build 82	Mm build 82: recent analysis	Mm build 100	Mm build 100: recent analysis	Mm build 120	Hs build 160
Y	*Guca1a*	guanylate cyclase activator 1a (retina)	*GUCA1A*	guanylate cyclase activator 1A (retina)					1	1
Y	*Guca1b*	guanylate cyclase activator 1B	*GUCA1B*	guanylate cyclase activator 1B (retina)				1	1	1
Y			*GUCA1C*	guanylate cyclase activator 1C						1
Y	*Gucy2e*	guanylate cyclase 2e							1	
Y	*Hes1*	hairy and enhancer of split 1 (Drosophila)	*HES1*	hairy and enhancer of split 1, (Drosophila)					1	1
Y	*Hes5*	hairy and enhancer of split 5 (Drosophila)			1		1		1	
N	*Hist1h2ai*	histone 2, H2ai					1		1	
N	*Hk1*	hexokinase 1							1	
N			*HMG20B*	high-mobility group 20B						1
N	*Hmgn2*	high mobility group nucleosomal binding domain 2					1		1	
N	*Hnrpk*	heterogeneous nuclear ribonucleoprotein K					1			
N	*Hspca*	heat shock protein 1, alpha					1			
N	*Hspa8*	heat shock protein 8					1			
N			*IGFBP5*	insulin-like growth factor binding protein 5						1
N	*Igf2bp1*	insulin-like growth factor 2, binding protein 1					1		1	
Y	*Igsf9*	immunoglobulin superfamily, member 9					1		1	
Y	*Impdh1*	inosine 5'-phosphate dehydrogenase 1					1			
Y	*Impg1*	interphotoreceptor matrix proteoglycan 1	*IMPG1*	interphotoreceptor matrix proteoglycan 1			1		1	1
N	*Ina*	internexin neuronal intermediate filament protein, alpha					1			
N			*ISL1*	ISL1 transcription factor, LIM/homeodomain, (islet-1)						1
N	*Itgb5*	integrin beta 5			1		1		1	
N	*Kcnab2*	potassium voltage-gated channel, shaker-related subfamily, beta member 2					1			
N	*Kcnb1*	potassium voltage gated channel, Shab-related subfamily, member 1	*KCNB1*	potassium voltage-gated channel, Shab-related subfamily, member 1		1	1		1	1
Y	*Kcnj14*	potassium inwardly-rectifying channel, subfamily J, member 14	*KCNJ14*	potasssium inwardly-rectifying channel, subfamily J, member 14		1		1	1	1
N			*KCNV2*	potasssium channel, subfamily V, member 2						1
Y	*Kera*	keratocan			1		1		1	
N	*Kif1A*	kinesin family member 1A					1			
N	*Kif1B*	kinesin family member 1B							1	
N	*Kpna2*	karyopherin (importin) alpha 2					1		1	
N			*KRT4*	keratin 4						1
N			*KRT5*	keratin 5 (epidermolysis bullosa simplex, Dowling-Meara/Kobner/Weber-Cockayne types)						1
Y			*KRT12*	keratin 12 (Meesmann corneal dystrophy)						1
N			*KRT18*	keratin 18						1
N			*KRT19*	keratin 19						1
N	*Lats2*	large tumor suppressor 2			1					
N	*Ldb1*	LIM domain binding			1		1		1	
Y	*Lhx2*	LIM homeobox protein 2			1		1		1	
Y	*Lim2*	lens intrinsic membrane protein 2	*LIM2*	lens intrinsic membrane protein 2, 19kDa				1	1	1
N	*Lmnb1*	lamin B1					1			
N	*Lox*	lysyl oxidase			1					
N			*LPHN1*	latrophilin 1						1
N			*LTBP3*	latent transforming growth factor beta binding protein 3						1
Y	*Mab21l1*	mab-21-like 1 (C. elegans)	*MAB21L1*	mab-21-like 1 (C. elegans)			1		1	1
Y	*Mab21l2*	mab-21-like 2 (C. elegans)					1		1	
N			*MAGI-3*	membrane-associated guanylate kinase-related (MAGI-3)						1
N	*Mak*	male germ cell-associated kinase	*MAK*	male germ cell-associated kinase			1		1	1
N	*Man2a2*	mannosidase 2, alpha2							1	
N			*MAPK3*	mitogen-activated protein kinase 3						1
N			*MAPK4*	mitogen-activated protein kinase 4						1
N	*Mapk8ip3*	mitogen-activated protein kinase 8 interacting protein 3							1	
N			*MAPT*	microtubule-associated protein tau						1
N			*MBD1*	methyl-CpG binding domain protein 1						1
N			*MFGE8*	milk fat globule-EGF factor 8 protein						1
Y			*MFRP*	membrane frizzled-related protein						1
N			*MGAT4B*	mannosyl (alpha-1,3-)-glycoprotein beta-1,4-N-acetylglucosaminyltransferase, isoenzyme B						1
Y	*Mip*	major intrinsic protein of eye lens fiber				1	1		1	
N	*Mlp*	MARCKS-like protein					1		1	
Y	*Mpp4*	membrane protein, palmitoylated 4 (MAGUK p55 subfamily member 4)					1		1	
N			*MSH6*	mutS homolog 6 (E. coli)						1
N	*Msi1h*	Musashi homolog 1(Drosophila)					1			
N	*Mtap1b*	microtubule-associated protein 1 B					1		1	
N	*Mtap6*	microtubule-associated protein 6					1		1	
N	*Mtch1*	mitochondria carrier homolog 1 (C.elegans)			1					
N			*MTMR9*	myotubularin related protein 9						1
N	*Mxi1*	Max interacting protein			1					
N			*MYH11*	myosin, heavy polypeptide 11, smooth muscle						1
N			*MYL9*	myosin, light polypeptide 9, regulatory						1
N	*Mylk*	myosin, light polypeptide kinase						1		
Y			*MYOC*	myocilin, trabecular meshwork inducible glucocorticoid response						1
N	*Nakap95*	neighbor of A-kinase anchoring protein 95	*NAKAP95*	neighbor of A-kinase anchoring protein 95		1				1
N	*Ncam1*	neural cell adhesion molecule 1							1	
N	*Ndr4*	N-myc downstream regulated 4						1	1	
N	*Nedd4*	neural precursor cell expressed, developmentally down-regulated gene 4			1		1		1	
N	*Nefl*	neurofilament, light polypeptide					1		1	
N	*Nes*	nestin			1					
Y	*Neurod1*	neurogenic differentiation 1					1		1	

Eye relevant	Mm Symbol	Mm Name	Hs Symbol	Hs Name	Mm build 82	Mm build 82: recent analysis	Mm build 100	Mm build 100: recent analysis	Mm build 120	Hs build 160
Y	*Neurod4*	neurogenic differentiation 4			1		1		1	
N	*Neurog2*	neurogenin 2							1	
N	*Nnat*	neuronatin					1		1	
N			*NP25*	neuronal protein						1
N			*NR1D1*	nuclear receptor subfamily 1, group D, member 1						1
Y	*Nr2e1*	nuclear receptor subfamily 2, group E, member 1						1	1	
Y	*Nr2e3*	nuclear receptor subfamily 2, group E, member 3			1		1		1	
Y	*Nrl*	neural retina leucine zipper gene	*NRL*	neural retina leucine zipper	1		1		1	1
N			*NRTN*	neurturin						1
N			*NRXN2*	neurexin 2						1
Y			*OCSP*	oculospanin						1
N	*Odc*	ornithine decarboxylase, structural			1	1	1		1	
N	*Olfm1*	olfactomedin 1			1					
Y	*Opn1mw*	opsin 1 (cone pigments), medium-wave-sensitive (color blindness, deutan)					1		1	
Y	*Opn1sw*	opsin 1 (cone pigments), short-wave-sensitive (color blindness, tritan)					1		1	
Y			*OPTC*	opticin						1
N			*OSAP*	ovary-specific acidic protein						1
N(Mm),Y(Hs)	*Osbp2*	oxysterol binding protein 2	*OSBP2*	oxysterol binding protein 2			1		1	1
Y	*Otx2*	orthodenticle homolog 2 (Drosophila)	*OTX2*	orthodenticle homolog 2 (Drosophila)				1	1	1
Y	*p*	pink-eyed dilution							1	
N			*P66beta*	transcription repressor p66 beta component of the MeCP1 complex						1
N	*Pacsin1*	protein kinase C and casein kinase substrate in neurons 1							1	
Y	*Pax6*	paired box gene 6	*PAX6*	paired box gene 6 (aniridia, keratitis)			1		1	1
N	*Pcbp2*	poly(rC) binding protein 2					1		1	
N	*Pcbp3*	poly(rC) binding protein 3					1		1	
N			*PCBP4*	poly(rC) binding protein 4						1
N			*PCDHB4*	protocadherin beta 4						1
N	*Pclo*	piccolo (presynaptic cytomatrix protein)					1			
N			*PCTAIRE2BP*	tudor repeat associator with PCTAIRE 2						1
N			*PCTK3*	PCTAIRE protein kinase 3						1
Y	*Pdc*	phosducin			1		1		1	
Y	*Pde6a*	phosphodiesterase 6A, cGMP-specific, rod, alpha	*PDE6A*	phosphodiesterase 6A, cGMP-specific, rod, alpha			1	1	1	1
Y	*Pde6b*	phosphodiesterase 6B, cGMP, rod receptor, beta polypeptide			1		1		1	
Y	*Pde6g*	phosphodiesterase 6G, cGMP-specific, rod, gamma	*PDE6G*	phosphodiesterase 6G, cGMP-specific, rod, gamma			1		1	1
Y			*PDE6H*	phosphodiesterase 6H, cGMP-specific, cone, gamma						1
N			*PEA15*	phosphoprotein enriched in astrocytes 15						1
N	*Pex2*	peroxin 2					1		1	
N	*Pfkfb2*	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 2							1	
N	*Pfkl*	phosphofructokinase, liver, B-type							1	
N	*Pfkp*	phosphofructokinase, platelet						1		
N	*Pftk1*	PFTAIRE protein kinase 1			1					
N	*Pgbpll*	pseudoautosomal GTP binding protein-like, like						1		
N			*PIPPIN*	ortholog of rat pippin						1
Y			*PITPNC1*	phosphatidylinositol transfer protein, cytoplasmic 1						1
N	*Pitpnm*	phosphatidylinositol membrane-associated					1		1	
Y			*PITX2*	paired-like homeodomain transcription factor 2						1
N	*Pkm2*	pyruvate kinase, muscle					1		1	
N			*PLA2G5*	phospholipase A2, group V						1
N	*Plagl1*	pleiomorphic adenoma gene-like 1							1	
Y			*PLEKHB1*	pleckstrin homology domain containing, family B (evectins) member 1						1
N	*Pold1*	polymerase (DNA directed), delta 1, catalytic subunit			1					
Y	*Pou4f2*	POU domain, class 4, transcription factor 2						1	1	
N			*PPP1R3F*	protein phosphatase 1, regulatory (inhibitor) subunit 3F						1
Y	*Prcad*	photoreceptor cadherin	-	MT-protocadherin				1	1	1
N			*PRDM8*	PR domain containing 8						1
N			*P-Rex1*	-						1
Y	*Prominl*	prominin-like 1							1	
Y	*Prox1*	prospero-related homeobox 1							1	
N	*Psmd3*	proteasome (prosome, macropain) 26S subunit, non-ATPase, 3			1					
N	*Ptbp1*	polypyrimidine tract binding protein 1			1					
N			*PTGDS*	prostaglandin D2 synthase 21kDa (brain)						1
N	*Ptma*	prothymosin alpha							1	
N	*Ptp4a3*	protein tyrosine phosphatase 4a3							1	
N			*PTRF*	polymerase I and transcript release factor						1
N	*Rab1*	RAB1, member RAS oncogene family			1		1		1	
N	*Rab5ef*	Rab5 exchange factor					1		1	
Y	*Rax*	retina and anterior neural fold homeobox	*RAX*	retina and anterior neural fold homeobox	1		1		2	1
Y	*Rbp3*	retinol binding protein 3, interstitial	*RBP3*	retinol binding protein 3, interstitial				1	1	1
Y			*RCV1*	recoverin						1
Y	*Rdh12*	retinol dehydrogenase 12	*RDH12*	retinol dehydrogenase 12 (all-trans and 9-cis)		1		1		1
Y	*Rds*	retinal degeneration, slow (retinitis pigmentosa 7)					1		1	
Y	*Rgr*	retinal G protein coupled receptor			1		1		1	
Y	*Rgs9bp*	regulator of G-protein signalling 9 binding protein							1	
N			*RGS19IP1*	regulator of G-protein signalling 19 interacting protein 1						1
Y	*Rho*	rhodopsin	*RHO*	rhodopsin (opsin 2, rod pigment) (retinitis pigmentosa 4, autosomal dominant)				1	1	1
Y	*Rhok*	rhodopsin kinase					1		1	

Eye relevant	Mm Symbol	Mm Name	Hs Symbol	Hs Name	Mm build 82	Mm build 82: recent analysis	Mm build 100	Mm build 100: recent analysis	Mm build 120	Hs build 160
Y	*Rlbp1*	retinaldehyde binding protein 1	*RLBP1*	retinaldehyde binding protein 1			1		1	1
N	*Rnf26*	ring finger protein 26				1				
N			*RNF110*	ring finger protein 110						1
Y	*Rom1*	rod outer segment membrane protein 1	*ROM1*	retinal outer segment membrane protein 1	1		1		1	1
N	*Ror1*	receptor tyrosine kinase-like orphan receptor 1					1		1	
Y	*Rorb*	RAR-related orphan receptor beta							2	
Y	*Rp1h*	retinitis pigmentosa 1 homolog (human)	*RP1*	retinitis pigmentosa 1 (autosomal dominant)			1		1	1
Y			*RP1L1*	retinitis pigmentosa 1-like 1						1
Y	*Rpe65*	retinal pigment epithelium	*RPE65*	retinal pigment epithelium-specific protein 65kDa					1	1
Y	*Rpgrip1*	retinitis pigmentosa GTPase regulator interacting protein 1					1		1	
N	*Rpl7a*	ribosomal protein L7a							1	
N	*Rpl13*	ribosomal protein L13							1	
Y	*Rps6ka2*	ribosomal protein S6 kinase, polypeptide 2							1	
N			*RRAD*	Ras-related associated with diabetes						1
Y	*Rrh*	retinal pigment epithelium derived rhodopsin homolog					1		1	
Y	*Rs1h*	retinoschisis 1 homolog (human)	*RS1*	retinoschisis (X-linked, juvenile) 1			1		1	1
Y			*RTBDN*	retbindin						1
N	*Rtn4*	reticulon 4							1	
N			*RTP801*	HIF-1 responsive RTP801						1
N			*RXRG*	retinoid X receptor, gamma						1
Y	*Sag*	retinal S-antigen	*SAG*	S-antigen; retina and pineal gland (arrestin)			1		1	1
N	*Sart1*	squamous cell carcinoma antigen recognized by T-cells 1			1				1	
N			*SAST*	syntrophin associated serine/threonine kinase						1
N	*Sav1*	salvador homolog 1 (Drosophila)				2	1		1	
N	*Scamp5*	secretory carrier membrane protein 5	*SCAMP5*	secretory carrier membrane protein 5					1	1
N	*Scd3*	stearoyl-coenzyme A desaturase 3						1		
N	*Scg2*	secretogranin II					1		1	
N	*Scgn*	secretagogin, EF-hand calcium binding protein							1	
N			*SDC3*	syndecan 3 (N-syndecan)						1
N			*SE20-4*	cutaneous T-cell lymphoma-associated tumor antigen se20-4						1
N			*SERPINA3*	serine (or cysteine) proteinase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 3						1
N			*SERPING1*	serine (or cysteine) proteinase inhibitor, clade G (C1 inhibitor), member 1, (angioedema, hereditary)						1
Y	*Sfrp2*	secreted frizzled-related sequence protein 2			1		1		1	
N	*Sfrs3*	splicing factor, arginine/serine-rich 3 (SRp20)					1			
Y	*Sfrs5*	splicing factor, arginine/serine-rich 5 (SRp40, HRS)							1	
Y	*Si*	silver	*SILV*	silver homolog (mouse)	1		1		1	1
N			*SIRT3*	sirtuin (silent mating type information regulation 2 homolog) 3 (S. cerevisiae)						1
Y	*Six3*	sine oculis-related homeobox 3 homolog (Drosophila)	*SIX3*	sine oculis homeobox homolog 3 (Drosophila)			1		1	1
Y	*Six6*	sine oculis-related homeobox 6 homolog (Drosophila)	*SIX6*	sine oculis homeobox homolog 6 (Drosophila)			1		1	1
Y	*Slc6a6*	solute carrier family 6 (neurotransmitter transporter, taurine), member 6					1		1	
N	*Slc6a9*	solute carrier family 6 (neurotransmitter transporter, glycine), member 9						1		
N			*SLC6A13*	solute carrier family 6 (neurotransmitter transporter, GABA), member 13						1
N			*SLC12A4*	solute carrier family 12 (potassium/chloride transporters), member 4						1
N			*SLC12A5*	solute carrier family 12, (potassium-chloride transporter) member 5						1
Y	*Slc16a8*	solute carrier family 16 (monocarboxylic acid transporters), member 8			1		1		1	
N	*Slc21a12*	solute carrier family 21 (organic anion transporter), member 12	*SLC21A12*	solute carrier family 21 (organic anion transporter), member 12		1		1	1	1
N			*SLC22A8*	solute carrier family 22 (organic anion transporter), member 8						1
N			*SLC23A1*	solute carrier family 23 (nucleobase transporters), member 1						1
Y	*Slc24a1*	solute carrier family 24 (sodium/potassium/calcium exchanger), member 1						1	1	
Y	*Slc38a1*	solute carrier family 38, member 1				1		1	1	
N	*Slc38a3*	solute carrier family 38, member 3	*SLC38A3*	solute carrier family 38, member 3					1	1
N	*Slit2*	Slit-like 2 (Drosophila)				1				
N	*Smarca4*	F related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 4						1	1	
N	*Smarce1*	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily e, member 1				1		1		
N	*Sox8*	SRY-box containing gene 8						1	1	
N	*Sox11*	SRY-box containing gene 11					1		1	
Y	*Spacrcan*	sialoprotein associated with cones and rods proteoglycan							1	
N			*SPG20*	spastic paraplegia 20, spartin (Troyer syndrome)						1
N			*SPTBN5*	spectrin, beta, non-erythrocytic 5						1
N	*Sqle*	squalene epoxidase							1	
Y	*Srfrs5*	splicing factor, arginine/serine-rich 5 (SRp40, HRS)					1			
N			*STK29*	serine/threonine kinase 29						1
N	*Stxbp1*	syntaxin binding protein 1					1		1	
N	*Sv2a*	synaptic vesicle glycoprotein 2 a					1			
Y			*SV2B*	synaptic vesicle glycoprotein 2B						1
N	*Syne1*	synaptic nuclear envelope 1							1	
Y	*Synpr*	synaptoporin				1		1		
N	*Syp*	synaptophysin					1			
N	*Syt1*	synaptotagmin 1					1		1	
N			*TBX2*	T-box 2						1
N	*Tcf4*	transcription factor 4			1					
N			*TERE1*	transitional epithelia response protein						1
N			*TF*	transferrin						1
Y			*TFAP2A*	transcription factor AP-2 alpha (activating enhancer binding protein 2 alpha)						1
Y			*TFAP2B*	transcription factor AP-2 beta (activating enhancer binding protein 2 beta)						1
Y			*TIMP3*	tissue inhibitor of metalloproteinase 3 (Sorsby fundus dystrophy, pseudoinflammatory)						1

Eye relevant	Mm Symbol	Mm Name	Hs Symbol	Hs Name	Mm build 82	Mm build 82: recent analysis	Mm build 100	Mm build 100: recent analysis	Mm build 120	Hs build 160
N	*Tm4sf2*	transmembrane 4 superfamily member 2					1			
N			*TMIE*	transmembrane inner ear protein						1
N			*TPD52L1*	tumor protein D52-like 1						1
N			*TRIM11*	tripartite motif-containing 11						1
N	*Trim28*	tripartite motif protein 28					1		1	
N	*Trpm1*	transient receptor potential cation channel, subfamily M, member 1							1	
N	*Trpm2*	transient receptor potential cation channel, subfamily M, member 2							1	
Y	*Tub*	tubby candidate gene					1			
N	*Tuba1*	tubulin, alpha 1					1		1	
N	*Tuba2*	tubulin, alpha 2					1			
N	*Tuba6*	tubulin, alpha 6							1	
N	*Tubb3*	tubulin, beta 3					1		1	
N	*Tubb4*	tubulin, beta 4					1			
N	*Tubb5*	tubulin, beta 5					1		1	
Y	*Tulp1*	tubby like protein 1	*TULP1*	tubby like protein 1			1		1	1
Y	*Tyr*	tyrosinase			1	1	1		1	
Y	*Tyrp1*	tyrosinase-related protein 1	*TYRP1*	tyrosinase-related protein 1	1					1
N	*Ubb*	ubiquitin B							1	
N	*Ubc*	ubiquitin C							1	
Y	*Unc119h*	unc1`9 homolog (C. elegans)	*UNC119*	unc-119 homolog (C. elegans)			1		1	1
N	*Usp11*	ubiquitin specific protease 11						1		
N	*Vamp2*	vesicle-associated membrane protein 2	*VAMP2*	vesicle-associated membrane protein 2 (synaptobrevin 2)			1		1	1
Y			*VMD2*	vitelliform macular dystrophy (Best disease, bestrophin)						1
N	*Vtn*	vitronectin							1	
N	*Wdr6*	WD repeat domain 6				1	1			
Y	*Wdr17*	WD repeat domain 17						1	1	
N			*WIF1*	WNT inhibitory factor 1						1
N			*ZNF151*	zinc finger protein 151 (pHZ-67)						1
N	*Zfp385*	zinc finger protein 385					1		1	
N			-	alpha-2 macroglobulin family protein VIP						1
N			-	netrin G2						1
N			-	NADH:ubiquinone oxidoreductase MLRQ subunit homolog						1
N	-	Na, K-ATPase alpha 1 isoform homolog (Salmo salar)							1	
Y	-	rootletin							1	

Appendix B

Known genes in rat DDD results

Rn Symbol	Rn Name
Aldh3a1	aldehyde dehydrogenase family 3, member A1
Btnl2	butyrophilin-like 2 (MHC class II associated)
C3	Complement component 3
Cbfb	core binding factor beta
Ccl5	chemokine (C-C motif) ligand 5
Cd74	CD74 antigen (invariant polpypeptide of major histocompatibility class II antigen-associated)
Clca2	chloride channel calcium activated 2
Cspg2	chondroitin sulfate proteoglycan 2 (versican)
Dcn	decorin
Emp1	Epithelial membrane protein 1
Gbp2	guanylate binding protein 2, interferon-inducible
Hadhb	hydroxyacyl-Coenzyme A dehydrogenase/3-ketoacyl-Coenzyme A thiolase/enoyl-Coenzyme A hydratase (trifunctional protein), beta subunit
Krt14	keratin 14
Krt2-5	keratin complex 2, basic, gene 5
Lum	lumican
Mgea5	meningioma expressed antigen 5 (hyaluronidase)
Mmp12	matrix metalloproteinase 12
Muc4	mucin 4
RT1Aw2	RT1 class Ib gene
S100a9	S100 calcium-binding protein A9 (calgranulin B)
Sdc4	syndecan 4
Slpi	secretory leukocyte protease inhibitor
unr	unr protein
Zfp361l	Zfp361l
-	MHC class antigen RT1.B-1 beta-chain
-	MHC class II RT.u-D-alpha chain